

AD-A215 830

# Profile of AMERICAN YOUTH

DEMOGRAPHIC  
INFLUENCES  
ON ASVAB TEST  
PERFORMANCE

By R. Darrell Bock and Elsie G. J. Moore  
National Opinion Research Center

DTIC  
ELECTE  
DEC 15 1989  
S B D



Office of Assistant Secretary of Defense  
(Manpower, Installations and Logistics)

FEBRUARY 1984

DISTRIBUTION STATEMENT A

Approved for public release  
Distribution Unlimited

89

12

16

01

1

THE PROFILE OF AMERICAN YOUTH

Demographic Influences On ASVAB  
Test Performance

R. Darrell Bock  
and  
Elsie G. J. Moore

National Opinion Research Center

Office of the Assistant Secretary of Defense  
(Manpower, Installations and Logistics)

February 1984

DTIC  
ELECTE  
DEC 15 1989  
S B D

DISTRIBUTION STATEMENT A

Approved for public release;  
Distribution Unlimited

# TABLE OF CONTENTS

PREFACE .....	ix
1. PROFILE OF AMERICAN YOUTH .....	1
1.1 The Profile Study .....	3
1.2 Approach to the Analysis.....	5
1.3 Approach to Interpretation .....	8
2. THE SAMPLE FOR THE PROFILE STUDY .....	9
2.1 Drawing the Sample .....	9
2.2 The Interviews.....	16
2.3 Recruitment of Subjects and Test Administration .....	16
2.4 Background Factors .....	20
2.5 Population Composition .....	29
3. THE TESTS.....	40
3.1 History of the ASVAB .....	40
3.2 The Rationale for Vocational Testing .....	43
3.3 Validity of the ASVAB .....	50
3.4 The Problem of Differential Predictive Validity.....	54
3.5 Description of ASVAB 8.....	75
3.6 Psychometric Properties of the ASVAB .....	89
3.7 Summary.....	95
4. METHODS OF ANALYSIS.....	97
4.1 Sampling Considerations .....	98
4.2 Data Properties .....	103
4.3 Tests of the Model .....	121
4.4 Interaction Displays .....	125
5. SOCIOCULTURAL EFFECTS .....	129
5.1 Sociocultural Group by Economic Status by Education Effects .....	131
5.2 Sociocultural Group by Region Effects .....	148
5.3 Sources of Sociocultural Group Differences .....	154
5.4 Summary and Conclusions.....	173
6. THE EFFECTS OF MOTHER'S EDUCATION AND RESPONDENT'S AGE AND EDUCATION .....	184
6.1 Mother's Education .....	184
6.2 Age and Education .....	198
6.3 Summary and Conclusion.....	204



By <u>See ADA111726</u>	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

# TABLE OF CONTENTS (Cont'd)

7.	SEX EFFECTS.....	207
7.1	Sex by Education Effects .....	208
7.2	Sex by Sociocultural Group by Economic Status .....	218
7.3	Sources of Sex Effects .....	230
7.4	Sex Differences in Spatial Ability .....	241
7.5	Summary and Conclusions.....	255
8.	SUMMARY AND CONCLUSIONS .....	259
8.1	The Sociocultural Groups Differ in ASVAB Test Performance According to Years of Education, Economic Status and Type of Test .....	260
8.2	Mother's Education and Years of Formal Schooling Are Strong Predictors of ASVAB Test Performance.....	273
8.3	Sex Differences Vary Most Strongly with the Type of Test.....	275
8.4	Differences in Test Performance of Young People at the Same Level of Education in Different Social and Economic Groups Show the Need for More Comprehensive Standards of Educational Attainment.....	280
	References .....	282



# LIST OF TABLES

Tables	Page
2.1-1 Cross-section sample youth located in screening, selected for baseyear interview, and completed Profile test .....	12
2.1-2 Supplemental sample youth located in screening, selected for baseyear interview, completed interview and completed Profile test .....	13
2.1-3 Military sample baseyear and Profile test completed cases .....	14
2.4-1 Profile subjects by year of birth .....	21
2.4-2 Profile subjects by highest grade completed .....	22
2.4-3 Profile subjects by sociocultural group .....	23
2.4-4 Profile subjects' self-reported origin or descent classified by sociocultural group .....	24
2.4-5 Profile subjects by economic status .....	26
2.4-6 Profile subjects by region of residence at age 14 .....	27
2.4-7 Profile subjects by mother's highest grade completed .....	28
2.5-1 Percent of NLS Year 1 sample in each sociocultural group and economic status, by region of residence at age 14 .....	30
2.5-2 Percent of NLS Year 1 sample completing each grade category by region of residence at age 14 .....	31
2.5-3 Percent of NLS Year 1 sample completing each grade category, by age group .....	32
2.5-4 Percent of noninstitutional United States population completing each grade category by age group .....	33
2.5-5 Percent of NLS Year 1 sample completing each grade category, by sociocultural group and economic status .....	33
2.5-6 Percent of NLS Year 1 sample completing each grade category by sociocultural group and age group .....	35
2.5-7 Percent of noninstitutional United States population completing each grade category by sociocultural group and age group .....	35
2.5-8 Percent of NLS Year 1 sample completing each grade category by sociocultural group and sex .....	37

# LIST OF TABLES (Cont'd)

2.5-9	Percent of NLS Year 1 sample whose mothers completed each grade category, by sociocultural group and economic status .....	38
2.5-10	Percent of NLS Year 1 sample completing each grade category, by grades completed by mother .....	39
3.2-1	Correlation of basic training predictors and criteria for a sample of Puerto Rican inductees .....	49
3.3-1	Single-sex validities of ASVAB composite scores as predictors of combined performance criterion for job groups defined during cross-validation .....	53
3.4-1	Sample sizes in aptitude area validation and racial studies....	63
3.4-2	Sample sizes, corrected and uncorrected predictor-criterion correlations, predictor composite, and final grade raw score means and standard deviations for eight Air Force training schools: reported by sociocultural group and sex.....	68
3.4-3	ASVAB-3 Predictor Composites: unselected raw score population means and standard deviations.....	70
3.6-1	Reliabilities of the ASVAB and the DAT .....	90
3.6-2	Pooled within-cell correlations (weighted) of ASVAB subtests .....	91
3.6-3	Likelihood ratio chi-square for fit of the factor analytic model assuming 1, 2, 3, 4 and 5 factors (N = 4,688) .....	92
3.6-4	Promax factor loadings and intercorrelations for the ASVAB test .....	93
4.1	Weighted and unweighted estimates of sociocultural group by economic status by education effects for the Word Knowledge test .....	100
4.2	Population composition for the age groups 15 through 22 estimated from the weighted analysis sample: highest grade completed by sociocultural group by economic status (N = 10,905) .....	106
4.3-1	Multivariate and univariate test statistics for effects of background factors: first analysis .....	122
4.3-2	Multivariate and univariate test statistics for effects of background factors: second analysis .....	123
4.3-3	Weighted marginal means of the ASVAB standardized scores for men and women and for Whites, Blacks, and Hispanics ....	128

# LIST OF ILLUSTRATIONS

Illustrations	Page
3.2-1 Employers' risks (percent of job failures among hired applicants) for various predictor validities ( $\rho$ = .2, .4, and .6) and selection ratios (percent of applicants hired) .....	45
3.4-1 Schematic selector vs. criterion regression lines for two groups.....	57
3.4-2 Comparative trends of course grades of Whites and Blacks for aptitude scores.....	64
3.4-3 Regressions for White and Black trainees of final course grade on selector composites for Air Force training schools.....	71
3.4-4 Regressions for male and female trainees of final course grade on selector composites for Air Force training schools.....	72
3.5-1a General Science and Arithmetic Reasoning: Selected item difficulties (higher scale numbers indicate greater difficulty) .....	77
3.5-1b Word Knowledge and Paragraph Comprehension: Selected item difficulties (higher scale numbers indicate greater difficulty) .....	78
3.5-1c Auto & Shop Information and Mathematics Knowledge: Selected item difficulties (higher scale numbers indicate greater difficulty) .....	79
3.5-1d Mechanical Comprehension and Electronics Information: Selected item difficulties (higher scale numbers indicates greater difficulty) .....	80
3.5-2 Sample problems: Numerical Operations and Coding Speed .....	81
4.2-1 Whites with some high school by sex.....	108
4.2-2 Whites with high school by sex.....	109
4.2-3 Whites with some college by sex.....	110
4.2-4 Blacks with some high school by sex.....	111
4.2-5 Blacks with high school by sex.....	112
4.2-6 Blacks with some college by sex.....	113

# LIST OF ILLUSTRATIONS (Cont'd)

4.2-7	Hispanics with some high school by sex.....	114
4.2-8	Hispanics with high school by sex.....	115
4.2-9	Hispanics with some college by sex.....	116
5.1a	Interaction of sociocultural group, economic status, and education (highest grade completed) .....	132
5.1b	Interaction of sociocultural group, economic status, and education (highest grade completed) .....	133
5.2	Interaction of sociocultural group and region of residence at age 14 .....	149
6.1-1	Effects of mother's education (mother's highest grade completed) .....	186
6.1-2	Interaction of mother's education (mother's highest grade completed) and sex .....	189
6.1-3	Interaction of sociocultural group and mother's education (mother's highest grade completed) .....	189
6.2	Interaction of age and education (highest grade completed) ...	202
7.1a	Interaction of sex and education (highest grade completed) ...	209
7.1b	Interaction of sex and education (highest grade completed) ...	210
7.2a	Interaction of sex, sociocultural group, and economic status .....	219
7.2b	Interaction of sex, sociocultural group, and economic status .....	220
7.4-1	Two items from the Thurstone and Thurstone Primary Mental Abilities Test (1947) .....	244
7.4-2	Identical Block Test (Stafford, 1961) .....	245
7.4-3	An item from the Vandenberg-Sheppard Mental Rotation Test ....	246

## PREFACE

In the Spring and Summer of 1980, the National Opinion Research Center (NORC) administered a battery of ten vocational tests to a national probability sample of nearly 12,000 young people, 15 to 23 years of age. The tests, the Armed Services Vocational Aptitude Battery (ASVAB), were the latest in a long series of aptitude and vocational tests developed by the Armed Services for purposes of selecting recruits and assigning them to military occupational specialties.

The purpose of this administration of the ASVAB by NORC was to obtain normative information on a representative contemporary civilian sample. The score distributions from this study were needed by the Armed Services to set the performance levels for accepting and assigning recruits.

In the belief that these and other results of this study would also be of interest to educators, psychologists, sociologists and the informed public, the staff of NORC invited us to examine and comment on the significance of the score distributions in light of background information obtained in interviews of these young people before and after testing. They asked us to draw upon our knowledge and experience in the fields of psychological measurement, human development, and minority group education in order to present the results in the context of previous research and issues in these fields. Their desire was to make the information in this virtually unique database available to a wider audience.

In examining these data the question that we had in mind from the beginning was, "What does the performance of these young people on these tests imply for their vocational prospects in the 1980s?" Our preliminary perusals of the data made evident a wide variation in test performance that inevitably would lead to marked advantages for some and disadvantages for others. We

foresaw, however, that any attempt to understand why this variation existed would present problems not resolveable on the basis of the data alone. To arrive at any reasonable interpretations, we expected to draw on research in many fields of behavioral science where the sources of individual differences in test performance have been investigated and discussed.

Much of our effort has therefore been devoted to review and synthesis of a literature that views human individual differences from mutually exclusive perspectives. Some authors only look for the effects of broad social structures and forces, others examine the influence of family, home and school, and still others concentrate on the biological foundations of behavior. Our vantage point from the field of human development affords a wider view enabling us to see more inclusively how all of these forces interact in the lives of American young people.

The present report is the product of our joint efforts. Most of the material in Chapters 3 and 4 giving background information about vocational testing in general and the ASVAB test in particular was prepared by the first author. The second author was primarily responsible for Chapters 5, 6 and 7 in which the effects of the various background factors on the test scores are examined. Exceptions are Section 7.4, devoted to sex differences in spatial visualizing ability, a topic of special interest to both authors and to which both of us contributed, and Section 5.4, which was the result of our joint discussions. Chapter 2, dealing with sampling issues, was prepared by the first author with the assistance of Harold McWilliams of NORC. The introductory and concluding chapters were the joint product of both authors.

Although we alone are responsible for our conclusions, we had the help of many others in all phases of this work. We acknowledge in particular the support and encouragement of A. J. Martin, former Director for Accession

Policy, Office of the Secretary of Defense; W.S. Sellman, Acting Director for Accession Policy, Office of the Secretary of Defense; Zahava D. Doering, Defense Manpower Data Center; B. K. Waters, Human Resources Research Organization; and Michael E. Borus, Project Director, National Longitudinal Survey of Youth Labor Force Behavior, Center for Human Resource Research, Ohio State University.

Staff of NORC directly responsible for sampling and data collection were Martin Frankel, Celia Homans, Mary Cay Burich, Shirley Knight, and Harold McWilliams. Harold McWilliams, Suzanne Erfurth, Susan Campbell and Steven Ingels were responsible for the final editing. Computer data processing was carried out by Robert Gibbons and Robert Mislevy. Michelle Zimowski, Mary Jo Petersen, and Ernest Froemel assisted in the review of the literature. Word processing was by Jean Hirsch, Anne Buckborough, Mary Okazaki, Chris Lonn, Irene Edwards, Mia Moore, Cassandra Britton, Karla Maze-Darby, Arturo Landeros, and Linda Budd. Linda Budd also prepared the graphs and figures.

## CHAPTER 1

### PROFILE OF AMERICAN YOUTH

According to recent Census data, approximately four million persons in the United States turn 18 years old each year. At this time in their lives they face major changes in situation and livelihood. Roughly 70 percent of the young men and women enter the labor market in some capacity, and nearly 10 percent enter the Armed Forces. Nearly 70 percent of them continue some type of full- or part-time higher education, often combining it with employment.<sup>1</sup> At the same time, about 20 percent marry and set up family households. In whatever position these young people find themselves, their opportunities frequently depend upon their performance on vocational and scholastic tests.

Many people applying for work in large business and industrial concerns must take vocational tests administered by the company. The usual practice of these employers is to consider further only those applicants who attain at least a minimally acceptable score, and among those, to select the applicants with the most favorable combinations of test scores, education, job interest, and experience. Alternatively, the tests may be given by an employment agency, which then prepares for the applicant a resume containing test scores, previous work history and other information of interest to prospective employers.

At various times the federal government has required persons applying for positions to take standardized examinations, such as the Civil Service Examinations. From time to time minimum acceptable scores for various federal government positions have been published.

---

<sup>1</sup>Statistical Abstract of the U.S. Bureau of the Census, Washington, DC, 1980.



Nearly all of the young people who choose to continue their education in colleges, universities, or specialized schools must submit results on standardized entrance tests with their applications. Approximately 1.25 million persons each year take the Scholastic Aptitude Test (SAT) prepared by the College Entrance Examination Board and administered by the Educational Testing Service.<sup>1</sup> In addition to or in place of the SAT, about one million young people take the entrance tests prepared and administered by the American College Testing Program (ACT).<sup>2</sup>

Volunteers for the Military Services must attain a certain minimum score on the Armed Services Vocational Aptitude Battery (ASVAB) as one of the requirements for entry into one of the Services and for assignment to training schools. This and other selection and classification tests are administered by various agencies of the Department of Defense. The Service Academies and the Coast Guard Academy have their own entrance test requirements, as do the specialized Federal training programs for various offices such as the Federal Aviation Agency, the Federal Bureau of Investigation, and the National Park Service.

This reliance on cognitive tests for personnel selection is a prominent feature of the meritocratic system that pervades contemporary society in the United States. In the meritocracy, evaluations of ability strongly determine position and reward. The system of universal primary and secondary education is the vehicle for nurturing ability, but school practices in the United States are too heterogeneous to allow school grades to be used exclusively in evaluating it. The standards on which school grades are based

-----

<sup>1</sup>College Entrance Examination Board, Annual Report, 1980-81 (New York).

<sup>2</sup>Personal communication with ACT, February, 1982.

vary so greatly among the states and between schools that external examinations such as those described above are needed in order to make fair comparisons among applicants. In countries with highly centralized school systems, such as France or the Soviet Union, school performance is a more dependable measure of ability, and external testing programs are less widely used. In the United States, however, these programs are necessary for the efficient and equitable allocation of employment and educational opportunities on the basis of merit. Their absence would work a hardship both on the organizations that would incur the costs of trying to train persons of insufficient work aptitudes, and on misassigned persons suffering the stress of failure on the job. The benefits of vocational testing are great enough to ensure that such programs will be a permanent feature of our society. For that reason, it is important that we have some assessment of their possible impact on the lives of American young people.

#### 1.1 The Profile Study

Considering the scope of standardized testing in the United States, it is surprising how little information on vocational test performance is available for a representative sample of young people. The score distributions of college-bound persons on the SAT and ACT tests are reported annually, but they are based on a specialized, self-selected sample of the population. Moreover, these tests are designed to predict success in college and not vocational success. The National Assessment of Educational Progress annually tests a national probability sample of in-school 17-year-olds, but persons who have already left school are not included, and the topics tested mostly reflect achievement in subject matter taught in secondary schools. The Current Population Surveys conducted by the U.S. Bureau of the Census collect self-reports of number of years of schooling completed, but do not attempt any actual testing of the respondents.

It is therefore of considerable importance that, in the summer of 1980, a broad-range battery of vocational aptitude tests was administered to a national probability sample of young people 15 to 23 years old.<sup>1</sup> The data were collected pursuant to a contract between the Department of Defense and the National Opinion Research Center (NORC), an affiliate of the University of Chicago. The purpose of the study was to develop current norms for the Armed Services Vocational Aptitude Battery.<sup>2</sup>

The data that resulted from this study have value far beyond the specialized technical objective of test standardization. They provide a unique opportunity to assess the vocational potential of the contemporary youth population and to compare the performance of persons of varying backgrounds and schooling. This study, called the "Profile of American Youth," yielded vocational test scores for 11,914 young men and women born between January 1, 1957 and December 31, 1964. The subjects were selected in 1978 for the Youth Cohort of the National Longitudinal Survey of Labor Force Behavior (NLS). From interviews in the spring of 1979 and 1980, background information on each subject was obtained that could be related to test performance.

One strength of the Profile study is that the data come from persons in all segments of society sampled by direct visits to households. They give a much broader picture of the vocational potential of American young people than tests administered to school children or to college-bound high school

-----  
<sup>1</sup>The 15-year-olds tested were part of the NLS sample and were to become 16 by the end of 1980.

<sup>2</sup>A series of reports has been prepared by NORC and the Department of Defense describing the Profile study. See Bock and Mislevy (1981), Department of Defense (1982), Frankel and McWilliams (1981), McWilliams (1982), McWilliams and Frankel (1982), and Sheatsley (1980).

seniors that have up to now been the main source of information about variation in ability in the population. Plans to include out-of-school young people in continuing testing programs such as the National Assessment of Educational Progress have not proved feasible because of the large costs involved.

Another strength of the Profile study is the comprehensiveness of the tests. The test battery, the ASVAB, is similar to the widely used Differential Aptitude Tests (DAT) and the General Aptitude Test Battery (GATB). It consists of ten separately timed and scored tests, the names of which briefly describe their content: General Science, Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, Mathematics Knowledge, Numerical Operations, Coding Speed, Mechanical Comprehension, Automotive-Shop Information, and Electronics Information. Reviewed in Chapter 3, these tests assess a range of knowledge and skills from general English-language reading and verbal skills through secondary-school mathematical knowledge and quantitative skills, to quite specific vocational knowledge in technical fields. Differences in profiles of attainment across the ten tests provide some of the more interesting results of the present study.

## 1.2 Approach to the Analysis

The main purpose of this report is the comparison of the test performance of certain groups of subjects classified by their background characteristics. The specific characteristics are age, sex, geographical region of residence at age 14, sociocultural group membership, economic status, highest grade completed in school, and mother's education. We interpret the results of these comparisons in the light of relevant literature from the fields of developmental psychology, differential psychology, educational and vocational testing, and occupational sociology.

Our treatment of these data is quite different from the multiple regression analyses used by sociologists such as Duncan, Jencks, and others to investigate variation in school achievement and vocational success (Duncan, et. al., 1972; Jencks, et. al., 1972, 1979; Solomon and Taubman, 1973). Unlike them, we are not attempting to express attainment as a quantitative index, nor are we assigning population-wide variation quantitatively to exogenous sources. Our approach is more qualitative. We look at a number of distinct response variables (the ASVAB test scores) in relation to the essentially qualitative background factors of the subjects. Viewed in this way, the data structure gives rise to a great number of estimated effects and interactions among factors that can be interpreted when the statistical tests indicate that interpretation is warranted. The statistics we will use in judging the significance of these effects and interactions take into account the fact that multiple contrasts and multiple variables are being examined simultaneously (see Chapter 4).

We find by these methods a number of significant and conceptually interesting interactions among the seven ways of classification examined in these data. Fortunately, no significant interactions involving more than three factors are observed. All effects in the study can therefore be displayed graphically as main effect and two- and three-way interaction plots for each ASVAB test. These plots, in the form used by Winer (1971), convey clearly both the qualitative and quantitative detail of the effects. We present the main results of our analyses in this form in Chapters 5, 6, and 7. These plots serve as points of departure for our discussion of the possible origins and implications of between-group differences in the data. In addition, because the response variables are scored in a standard form, certain comparisons between variables are possible. In many cases, it is the

differing impact of the sample classes on different tests that is the clearest clue as to the origin of the effects.

Inspection of the plots in Chapters 5, 6, and 7 reveals substantial differences in the average scores of subjects classified by various background characteristics. In some cases these differences exceed 1 standard deviation of the score distributions in the national probability sample. Among the most strongly interacting effects are those of sociocultural group, years of education, and economic status. Although the factors in such an interaction have no special order of precedence, we take the sociocultural group classification as a point of reference when discussing the specific effects of education and economic status. This discussion appears in Chapter 5. We continue our substantive discussion of the study with maternal effects in Chapter 6. This chapter also deals with the direct effects of the years of education of the subjects themselves, which of course are also large. Many of the subjects are, however, still in the midst of their high school or college education, and it is difficult to separate the effects of schooling from general developmental effects due to increasing age. We make an attempt to separate the two influences in Chapter 6 by comparing subjects in successive age groups who have different numbers of years of schooling. Some interesting differences appear in these comparisons, but their interpretation remains somewhat equivocal.

Finally, we discuss in Chapter 7 the attainment profile differences that characterize male and female subjects and the interaction of these differences with other factors such as years of education. Recent research activity in the area of sex differences in cognition provides a rich background for interpretation of the substantial sex effects found in the Profile data.

The remainder of our discussion is devoted to the methodological basis of the study and to summary and conclusions. The method of obtaining the sample of subjects is described in Chapter 2, and the content and characteristics of the tests administered to the subjects are presented in Chapter 3. The statistical methods employed in the analysis are outlined in Chapter 4. Finally, Chapter 8 summarizes all findings and brings together the inferences and conclusions we have drawn in the course of discussing particular findings.

### 1.3 Approach to Interpretation

The results of the Profile study presented here are descriptive of differences in average test performance among certain groups of young people in the United States in the summer of 1980. By themselves, they do not speak to the causes or sources of these differences; they exist instead as facts demanding explanation. In seeking to account for them plausibly we turn to the research literature on individual and group differences in test performance, educational attainment, and vocational success. No single literature covers all aspects of these topics. Relevant studies are found in the psychology of individual differences, in factor analyses of psychological tests, in the sociology of class, ethnic and cultural norms, in educational research on school effects in achievement, and even in historical research concerning the origins of the present populations of the United States.

Taking a broadly developmental point of view, we attempt to bring as much of this literature as possible to bear on the problem of identifying and interpreting the sources of the differences that appear in the data. At the same time, we address our own professional concerns in vocational and educational testing and counseling to suggest what implications the data may have for the future preparation of young people who must find their vocations and make places for themselves in the American meritocracy.

## CHAPTER 2

### THE SAMPLE FOR THE PROFILE STUDY<sup>1</sup>

Most of what we know about the abilities of young people is based on test scores of students in high schools and colleges. It is relatively easy to test subjects in these settings, and national and even international data are available for many types of attainment variables. Obtaining test results of a nationally representative sample of youth and young adults, the goal of the "Profile of American Youth," was a vastly more difficult undertaking. Because there was no institutional route to obtaining such subjects, they had to be located by interviewing them or their parents in their homes. Once identified, the subjects had to be persuaded to come to testing sessions and spend several hours in the arduous task of taking tests. For the Profile study, the sampling, interviewing and testing were carried out by the National Opinion Research Center (NORC), an affiliate of the University of Chicago.

#### 2.1 Drawing the Sample

The sample was initially drawn for the current National Longitudinal Survey of Youth Labor Force Behavior (NLS).<sup>2</sup> Subjects in the sample represent a cohort of persons born between January 1, 1957 and December 31, 1964. The NLS study design called for annual personal interviews beginning in 1979 and continuing for six years. The interviews concern the participants' educational

-----

<sup>1</sup>For additional information on the sample see Frankel and McWilliams (1981) and McWilliams and Frankel (1982).

<sup>2</sup>The NLS was initially funded by the Department of Labor under authority of the Comprehensive Employment and Training Act. Supplementary funding has been provided by the Office of the Assistant Secretary of Defense (Manpower, Installations and Logistics) and the Military Services. The prime contractor for the NLS is the Center for Human Resource Research of the Ohio State University. The National Opinion Research Center, University of Chicago, is the subcontractor for data collection.



and vocational experiences, their income, their state of health, and other demographic and attitudinal characteristics. At the time of the Profile testing, the subjects had completed two of the NLS interviews (1979 and 1980). These interviews provided background information that could be related to the subjects' scores on the ASVAB.

The NLS/Profile sample consists of three independent probability samples: (1) a cross-section sample designed to represent the noninstitutional civilian segment of American young people 14 to 21 years old as of January 1, 1979, in their proper population proportions; (2) a supplemental oversample of civilian Hispanic, Black, and economically disadvantaged non-Hispanic, non-Black (poor White) youth in the same age range; and (3) a military sample designed to represent youth aged 17 to 21 as of January 1, 1979 who were serving in the military as of September 30, 1978. The data from all three of these samples are used in the present study.

The two civilian samples were selected by standard area probability methods.<sup>1</sup> Stratification was introduced at several stages to insure a sufficient number of individuals from both urban and rural areas and from all geographical Census Divisions. Special procedures were employed to insure the inclusion of college students and persons living in group quarters.

Selection of the civilian samples involved several steps:

- random probability selection of 202 geographically dispersed county-size areas (primary sampling units)
- selection within each primary sampling unit of nine census blocks or enumeration districts
- listing the addresses of dwelling units within each selected census block or enumeration district

-----  
<sup>1</sup>These methods involve the random selection of progressively smaller units (counties, census blocks or enumeration districts, dwelling units, and individuals) to obtain a representative national probability sample.

- probability selection of a sample of the listed dwelling units
- completion of a screening interview with an adult in each selected dwelling unit to locate youth eligible for the survey.

In the fall of 1978, NORC interviewers screened approximately 81,000 households and located some 14,000 eligible youth. All eligible youth located were scheduled for a baseyear (1979) interview.<sup>1</sup> Tables 2.1-1 and 2.1-2 show the number of individuals located in screening, selected for a baseyear interview, interviewed, and tested in the cross-section and in the supplemental sample.

The military sample was selected from personnel files provided by the Department of Defense. Before selection, the files were stratified by the four Military Services and by five world-wide geographic regions. Females were oversampled at a rate six times that of males in order to produce approximately equal numbers of males and females in the final military sample. Table 2.1-3 shows the distribution of NLS baseyear respondents and Profile subjects by sex and Military Service.

Clustering of the Sample. The selection processes just described produced a geographically clustered (rather than randomly distributed) national probability sample. Clustering reduces the costs of contacting and interviewing the sample members in large-scale surveys because fieldworkers can be dispatched to a smaller number of sites than would be the case if the geographic distribution of the sample were random. Despite the clustering,

-----

<sup>1</sup>In screening for the supplemental sample, the number of Hispanic and Black youth located in the households selected for screening exceeded the number required in the sample design. Therefore, it was necessary to select a subsample of these individuals for baseyear interviewing. Table 2.1-2 shows the 714 cases deleted by subsampling. The remaining Hispanic and Black cases received greater caseweights to compensate for the deleted cases.

TABLE 2.1-1

CROSS-SECTION SAMPLE YOUTH LOCATED IN SCREENING, SELECTED FOR BASEYEAR INTERVIEW, COMPLETED INTERVIEW AND COMPLETED PROFILE TEST

Design Cohort	Located in Screening	Out of Scope (Ineligible)	Selected for Interview	Completed Baseyear Interview	Completed Profile Test
<b>Males:</b>					
Hispanic .....	249	5	244	216	195
Non-Hispanic Black	386	7	379	347	332
Economically dis- advantaged non- Hispanic, non-Black	189	5	184	203 <sup>a</sup>	188
Other .....	2,599	39	2,560	2,238	2,107
Total .....	3,423	56	3,367	3,004	2,822
<b>Females:</b>					
Hispanic .....	248		248	228	218
Non-Hispanic Black	451	13	438	404	388
Economically dis- advantaged non- Hispanic, non-Black	180		180	198 <sup>a</sup>	186
Other .....	2,620	41	2,579	2,277	2,152
Total .....	3,499	54	3,445	3,107	2,944
TOTAL .....	6,922 <sup>b</sup>	110	6,812	6,111 <sup>c</sup>	5,766 <sup>d</sup>

<sup>a</sup>A number of individuals who had not been identified as economically disadvantaged at the time of screening and selection for interview were reclassified after their interview. This accounts for the apparent discrepancy between the number selected and the number interviewed in this design cohort.

<sup>b</sup>Screening interviews were completed in 91.2 percent of the occupied dwelling units selected for screening.

<sup>c</sup>Interviews were completed with 89.7 percent of those selected for baseyear interviews.

<sup>d</sup>Profile tests were completed by 94.4 percent of those who completed a baseyear interview.

TABLE 2.1-2

SUPPLEMENTAL SAMPLE YOUTH LOCATED IN SCREENING, SELECTED FOR BASEYEAR INTERVIEW, COMPLETED INTERVIEW AND COMPLETED PROFILE TEST

Design Cohort	Located in Screening	Out of Scope (Ineligible)	Deleted by Subsampling	Selected for Interview	Completed Baseyear Interview	Completed Profile Test
<b>Males:</b>						
Hispanic.....	1,015	26	161	828	730	668
Non-Hispanic Black.....	1,318	42	50	1,226	1,096	1,043
Economically disadvantaged						
non-Hispanic non-Black.....	887	23	1	858	744	697
Other.....	--	--	--	--	--	--
Total.....	3,220	96	212	2,912	2,570	2,408
<b>Females:</b>						
Hispanic.....	1,060	16	205	839	750	695
Non-Hispanic Black.....	1,502	31	298	1,173	1,076	1,041
Economically disadvantaged						
non-Hispanic non-Black.....	1,073	28	--	1,045	899	846
Other.....	--	--	--	--	--	--
Total.....	3,625	75	503	3,057	2,725	2,582
TOTAL.....	6,855 <sup>a</sup>	171	715	5,969	5,295 <sup>b</sup>	4,990 <sup>c</sup>

<sup>a</sup>Screening interviews were completed in 91.9 percent of the occupied dwelling units selected for screening.

<sup>b</sup>Interviews were completed with 88.7 percent of those selected for baseyear interviews.

<sup>c</sup>Profile tests were completed by 94.2 percent of those who completed a baseyear interview.

TABLE 2.1-3

MILITARY SAMPLE BASEYEAR INTERVIEW AND  
PROFILE TEST COMPLETED CASES

Service	Completed Interview	Completed Profile Test
<b>Males:</b>		
Army.....	352	320
Navy.....	212	182
Air Force.....	163	151
Marine Corps.....	96	85
Total.....	823	738
<b>Females:</b>		
Army.....	224	208
Navy.....	68	59
Air Force.....	131	121
Marine Corps.....	34	32
Total.....	457	420
<b>TOTAL</b> .....	<b>1,280<sup>a</sup></b>	<b>1,158<sup>b</sup></b>

<sup>a</sup>Interviews were completed with 71.5 percent of those selected for baseyear interviews.

<sup>b</sup>Profile tests were completed by 90.5 percent of those who completed a baseyear interview.

every individual in the target age range in every part of the country has a chance of entering the sample. For a variety of reasons, however, the selected individuals have differential probabilities (i.e. unequal chances) of selection. To compensate for this a system of weighting is employed whereby the sample is once again made representative of the target population. (See Frankel and McWilliams, 1981, for a detailed description of the weighting scheme.)

Although clustering does not bias the sample, it does affect how the data must be treated to assess the statistical significance of results and to construct confidence intervals. Cluster samples show less internal sampling variability but possess greater standard errors than simple random samples of the same size. However, because the amount of reduction in variance can be estimated, the effect of clustering can be allowed for in statistical inference.

One of the more potent sources of the clustering effect is the selection of all persons in the eligible age range within each household. Of the 12,686 respondents to the Year 1 NLS interview, 4,197 came from households with one respondent, 4,304 from those with two, 2,154 with three, 948 with four, 305 with five, 60 with six, and 718 from other size households. Some of those from multiple-respondent households are siblings, some are spouses, some unrelated. Nevertheless, to the extent that they share a common environment, common heredity, or exhibit correlated mate selection, there will exist within-household (intraclass) correlation that can reduce the effective sample size.

In the present analysis, however, the young people who took the ASVAB tests are assigned to different background subgroups if they are not of the same sex or are at different education levels according to highest grade completed at the time of interview. The probability that two members of the same household will appear in the same subclasses of the analysis (and, by virtue of their similarity in test performance, bias downward the estimated error variance) is therefore considerably reduced. For this reason, we

believe that the cluster effect has only a minor impact on our analysis. We discuss the correction for clustering in Section 4.1.

## 2.2 The Interviews

In the spring of 1979, NORC conducted the first round of interviews for the NLS. The interviews covered a variety of background factors that influence the vocational prospects of young people entering the labor market. These interviews provided the demographic information used in the present study.

Extraordinary measures were taken to locate and gain the cooperation of the young people for the initial interview. As can be seen in Tables 2.1-1, 2.1-2, and 2.1-3, this effort met with considerable success: of the 6,812 subjects selected for the cross-section sample, 89.7 percent completed first year interviews. The interview completion rate for the supplemental sample of 5,969 Blacks, lower-income Whites and Hispanics was 88.7 percent. In the sample of 1,793 persons selected from the military (with oversampling of women), the completion rate fell to 71.5 percent, largely because of difficulty in reaching service personnel on remote assignment. Because the military sample constitutes only a small fraction of the total sample, however, this lower completion rate has little effect on the broad comparisons made in this study. For these purposes, the overall interview completion rate of 87.0 percent--higher than is typical in large-scale surveys other than the Census--is quite satisfactory.

## 2.3 Recruitment of Subjects and Test Administration

Because this was the first attempt to administer a battery of tests to a large-scale national probability sample, there was little experience to suggest how difficult it might be to convince subjects, on an entirely voluntary basis, to devote three and one-half hours of their time to taking the ten ASVAB tests. Obviously, this would be a far greater imposition on the subject

than the hour or so for the NLS survey interview. Moreover, to simulate as closely as possible the standard conditions under which the ASVAB was administered to military applicants, the battery was to be administered to groups of about ten subjects meeting in the same room at the same time, thus requiring the young people to travel at least a minimal distance to a testing site and increasing the total time they must invest.

The project staff realized that many subjects would participate in the study because they would be interested in receiving their test results or might consider taking the ASVAB good practice for other tests. Others, however, would probably require a monetary incentive to participate. A pretest conducted in the spring of 1980 indicated that a satisfactory rate of participation could be obtained by paying an honorarium of \$50. Therefore, because a high response rate was critical to the validity of the overall results, several incentives were used. Participants received copies of their test results, information to interpret the scores, a brochure containing vocational and educational information, and an honorarium of the same amount as proved effective in the pretest.

The participants were first informed of the testing project in a thank-you letter sent in May 1980 after the completion of the 1980 NLS (Year 2) interview. In June of that year, a formal request for participation was sent to the young people or to the parents of those under 18 years of age. This letter contained an attractive brochure announcing the "Profile of American Youth" and conveying the national importance of the project. The letter explained that participants would receive an honorarium of \$50, an informative brochure, and a report of their own scores, but that their individual scores would not be released to anyone else. The letter indicated that an NORC interviewer would



contact them soon to answer any questions they might have and to arrange a convenient time and place for the testing.<sup>1</sup>

Responsibility for contacting subjects was delegated to some two hundred fieldworker teams consisting of two persons, one of whom acted as examiner during the testing sessions and the other as the test proctor. Both workers participated in the job of locating and encouraging members of the NLS sample to take the tests. On average, each team was responsible for approximately sixty participants, the names, addresses, and telephone numbers of whom were supplied by NORC. The first action of the fieldworker was to send a letter to each participant on his or her list or, in the case of subjects less than 18 years of age, to the parent or guardian, informing them that they could expect a call from the interviewer some time in the next week. In the majority of cases, the youth could be reached by telephone, but in some instances the interviewers had to call on people at their places of residence. Participants who had been identified in the 1979 interview as primarily Spanish-speaking were sent letters in Spanish and contacted by bilingual interviewers.

Inevitably, there were some individuals on the roster of 12,686 participants in the 1979 interview who could not be located and tested. But as Tables 2.1-1, 2.1-2, and 2.1-3 testify, the numbers were surprisingly small. Of 12,686 to be tested, only 349 could not be located. They had moved and left no forwarding address or were totally inaccessible. Another 16 were deceased, and 61 of those who were located could not be tested for reasons of health, because they were incarcerated, or for miscellaneous other reasons.

---

<sup>1</sup>Copies of the letters to subjects and their parents and of the brochures may be found in the Appendix, Part B, of the Profile of American Youth: Field Report, McWilliams, 1980.

Finally, another 346 were located but refused to take the tests. The fieldworkers made substantial efforts to convince them that they should cooperate, but undue pressure was not exerted for fear of alienating them from cooperation in the remaining interviews in the longitudinal study.

In all, 11,914 subjects took the ASVAB. Most were tested in groups of about ten. The testing sites arranged by the fieldworkers typically were meeting rooms in hotels or motels, schools, libraries, or government buildings. About 700 subjects were tested individually. Whatever the site, the testing conditions and the times allowed for each test (measured by stop-watch), were exactly those prescribed for Military Entrance Processing Station (MEPS) administration of the tests.<sup>1</sup>

A few participants, thirty-six in all, had some sort of physical impairment of vision, motor control, etc., that made it possible for them to complete the test only with some modification in the prescribed conditions. For purposes of the present analysis, these subjects are excluded because the difficulty of the test is altered and the scores do not have comparable meaning when standard conditions are not adhered to. This leaves 11,878 subjects who completed the test under standard conditions and who constitute the sample for purposes of the present study. In all, the test completion rate of 93.63 percent is excellent, considering the exigencies of testing a population-wide sample.

-----  
<sup>1</sup>There was one minor departure from standard procedure: for eighty-five respondents the instructions for marking the answer sheet were translated into Spanish as an aid to those respondents with limited ability to read English. Individual items on the test and the instructions prior to each of the ten test sections, including the practice items, were not translated. Insofar as any of the tests required verbal ability, it was ability in the English language that was measured.

## 2.4 Background Factors

Although the 1979 and 1980 interviews collected data on a wide range of demographic and other background characteristics, the present study is concerned primarily with those characteristics that previous research suggested would be most associated with score variation on the ten ASVAB tests. These are: age, sex, education level, sociocultural group, economic status, region of residence at age 14, and mother's education. The classes within each of these seven factors are as follows:

1. Age: (15-17), (18-19), (20-22).
2. Sex: Male, female.
3. Education Level: [Highest grade completed] (0-8), (9-11), (12), (13 plus).
4. Sociocultural group: White, Black, Hispanic.
5. Economic status: Poor (at or below the official poverty level), nonpoor.
6. Region of residence at age 14: Northeast, Southeast, Midwest, West.
7. Mother's education: [Highest grade completed] (0-8), (9-11), (12), (13-15), (16 plus).

To give some idea of the kinds of responses on which these classifications are based, we present in this section tables showing frequency counts of the number of subjects in the categories of each background factor. These frequency counts are unweighted and cannot be used to estimate the distribution of these characteristics in the population. They reveal, however, the qualitative variety of the responses and the minor extent of missing or unusable information. In addition, we describe the definition and construction of each background factor.

Age. The analyses reported here employ year of birth as the age variable. Date of birth of the subjects was ascertained during the screening

interview conducted for purposes of sampling the target cohort and was confirmed or revised on the basis of later interviews with the young people themselves. Table 2.4-1 shows the distribution of the 11,914 subjects by year of birth. Since the test battery was administered during the summer of 1980, about half of the subjects in each cohort were older than the January 1980 age indicated in Table 2.4-1.

TABLE 2.4-1  
PROFILE SUBJECTS BY YEAR OF BIRTH

Year of Birth	1980 Age <sup>a</sup>	Number of Subjects
1957	22	1,545
1958	21	1,542
1959	20	1,597
1960	19	1,534
1961	18	1,447
1962	17	1,538
1963	16	1,488
1964	15	1,223 <sup>b</sup>
TOTAL		11,914

<sup>a</sup>Age as of January 1, 1980.

<sup>b</sup>Sample was underrepresented in youth born between October 1964 and December 1964 due to a programming error during the sample selection process.

Sex. Of the 11,914 subjects tested, 5,969 were male and 5,945 were female.

Education Level. Information on highest grade completed was obtained in both the 1979 and 1980 interviews. Subjects interviewed in 1980 are classified according to their response to the question, "What is the highest grade or year of regular school that you have completed or gotten credit for?" Those subjects not interviewed in 1980 are classified according to

their 1979 response to the same question. The distribution of subjects by highest grade completed is shown in Table 2.4-2.

TABLE 2.4-2  
PROFILE SUBJECTS BY HIGHEST GRADE COMPLETED

Highest Grade Completed	Number of Subjects
0	2
1	7
2	2
3	15
4	15
5	16
6	72
7	170
8	700
9	1,783
10	1,864
11	1,756
12	3,606
13	829
14	539
15	317
16	208
17	3
No information	10
TOTAL	11,914

Sociocultural Group. Among the many possible ways to disaggregate the United States population we have chosen to utilize a scheme that classifies the population in three broad sociocultural groups: White and Other, Black, and Hispanic. These same broad groupings are widely used in social science research and in government reports such as the Current Population Reports published by the U.S. Bureau of the Census.

Table 2.4-3 shows the distribution of Profile subjects across the three sociocultural groups. The sociocultural group categories are based on classes in the sampling design determined from the screening interview conducted with an adult member of the respondent's household. After the screening fieldwork was completed, individuals were assigned to one of the

three sociocultural groups based on three types of information obtained during the screening interview: (1) race (White, Black, or Other) as observed by the interviewer; (2) origin or descent as indicated in response to the question: "What is (person's) origin or descent?"; (3) whether Spanish was spoken in the home as a child. On the basis of this information, NORC's sampling department assigned individuals in the eligible age range in the screened households to one of three sociocultural groups. Individuals were assigned to the Hispanic group if their origin or descent had been recorded as Mexican-American, Chicano, Mexican, Mexicano, Cuban, Cubano, Puerto Rican, Puertorriqueno, Boricua, Latino, Other Latin American, Hispano, or of Spanish descent. In addition, individuals whose descent was other than those listed above were assigned to the Hispanic group if they met all of the following criteria: (1) they were not of Filipino or Portuguese descent, (2) Spanish was spoken in the home as a child, and (3) the family's surname appeared on the U.S. Census Bureau's list of Spanish surnames.

TABLE 2.4-3

PROFILE SUBJECTS BY SOCIOCULTURAL GROUP	
Sociocultural Group	Number of Subjects
White	7,043
Black	3,028
Hispanic	1,843
TOTAL	11,914

Individuals were classified as Black if the interviewer had coded their race as Black and their origin or descent was not Hispanic. In addition, individuals who identified their own origin or descent as Black, Negro, or Afro-American were classified as Black regardless of the race recorded by the interviewer. Thus, self-identification took precedence over the interviewer's classification.

All individuals not classified as Hispanic or Black were classified as members of the White sociocultural group.

The three sociocultural groups, in fact, contain considerable ethnic diversity. Thus, in Table 2.4-4 we show the unweighted frequency counts of the responses to a question in the 1979 Year 1 NLS interview: "What is your origin or descent?" These counts, based on the self-identifications of the Year 1 respondents themselves, give some indication of the heterogeneity of the three broad groupings.

TABLE 2.4-4

PROFILE SUBJECTS' SELF-REPORTED ORIGIN OR DESCENT  
CLASSIFIED BY SOCIOCULTURAL GROUP

Origin or Descent	Number of Subjects
White:	
Chinese	24
English	1,424
Filipino	41
French	274
German	1,268
Greek	27
Hawaiian	19
Native American	577
Asian-Indian	18
Irish	873
Italian	451
Japanese, Korean, Vietnamese	22
Polish	211
Portuguese	92
Russian	40
Scottish	116
Welsh	34
Other	725
Black	2,893
Hispanic:	
Cuban	96
Mexican, Mexican-American, Chicano	1,104
Puerto Rican	287
Other Latino	110
Other Spanish	86
Unclassifiable:	
American	703
None	128
Refused, Don't Know, or Missing	271
TOTAL	11,914

Our intent in using this classification of sociocultural group is not to distinguish races in an evolutionary sense but, rather, to identify the origins of the subjects in the broad subcultures that exist in the United States at the present time. Our decision to use these three broad groupings was based on the fact that most young people in the United States are nurtured in one or another of these three broad cultures; few belong simultaneously to more than one even when their parents are the products of different cultures. The definition of the culture is elusive, but its collective influence through family, peers, and community is unmistakable. It heavily determines manners, dress, spoken language, preferences in reading, music, entertainment and sports, personal values, and, most important for our purposes, educational and vocational aspirations. Among other things, it is the effects of these influences that we expect to see in the test scores of our subjects when we examine them by sociocultural group in Chapter 5.

Economic Status. Economic status, as measured by the income available to a family, exerts a pervasive and powerful influence on vocational prospects in most societies, including the United States. Economic status cuts across and acts within each of the broad sociocultural groups defined above. Therefore, in the analysis that follows, it is often used in combination with the sociocultural group factor to differentiate between the higher- and lower-income segments of the three sociocultural groups.

Subjects are classified on the economic status dimension as either Poor or Nonpoor based on definitions formulated by the Office of Management and Budget (OMB). These definitions of poverty take into account family size and urban or rural residence as well as total family income. For nonfarm families living in the continental United States, poverty is defined to be a net 1979 family income less than or equal to \$3,770 plus \$1,230 times one less



than the number of persons in the family unit. For farm families the formula is \$3,220 plus \$1,040 times one less than the number of persons in the family unit. For a nonfarm family of four, this translates into a definition of poverty level as a net family income of \$7,460 or less per year. By most people's standards this is a stringent definition of poverty, and as we will see in Section 2.5, only a small percent of the population falls within it.

Table 2.4-5 shows the distribution of Profile subjects by economic status based on their 1979 family income or, if they were not interviewed in 1980, based on their 1978 family income (using 1978 poverty definitions). Unfortunately this definition of poverty-line households does not distinguish between young people living with their parents and those living independently and thus classified as heads of households. The latter include, of course, students who are supporting themselves while in college or graduate school, or in other types of occupational training. Such persons are not genuinely poor; within a few years their earnings will be above, often far above, the poverty level.

TABLE 2.4-5  
PROFILE SUBJECTS BY ECONOMIC STATUS (UNWEIGHTED)

Economic Status	Number of Subjects
Poor	2,779
Nonpoor	8,813
No information	322
TOTAL	11,914

Note: Economic status is based on OMB criteria using 1979 and 1980 NLS interview data on family size and net family income.

Residence at Age 14. For purposes of the present analysis, subjects are classified according to the region of the country in which they were living at age 14. Residence at age 14 rather than residence at the time of

testing or interview is used because it is assumed to be a better indicator of where participants spent their formative years. It is possible, of course, that some subjects spent only a few years or even months in the region where they were living at age 14. Internal evidence (presented in Chapter 6) indicates, however, that residence at age 14 is an informative classification.

The assignment of subjects to regions is based on their responses to a question in the 1979 interview which asked where they were living on their 14th birthday. Subjects were classified first by state or territory of residence. The fifty states were then grouped into Northeast, Southeast, Midwest, and West regions according to the definition used by the National Assessment of Educational Progress. The few subjects residing in outlying territories of the United States were assigned to one of the four regions as described in the note to Table 2.4-6. The number of subjects in each region is shown in Table 2.4-6.

TABLE 2.4-6  
PROFILE SUBJECTS BY REGION OF RESIDENCE AT AGE 14

Region	Number of Subjects
Northeast	2,595
Southeast	3,018
Midwest	3,019
West	2,934
Elsewhere	265
No information	83
TOTAL	11,914

NOTE: Northeast region includes the following states and territories: CT, DE, DC, ME, MD, MA, NH, NJ, NY, PA, RI, VT, and Canal Zone, Puerto Rico, and Virgin Islands.

Southeast region includes the following states: AL, AR, FL, GA, KY, LA, MS, NC, SC, TN, VA, and WV.

Midwest region includes the following states: IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, and WI.

West region includes the following states and territories: AK, AZ, CA, CO, HI, ID, MT, NV, NM, OK, OR, TX, UT, WA, WY and Pacific Trust Territory and Other Pacific Islands.

Mother's Education. In order to have more information about family background than is carried by the subpopulation classification, we have also included the subject's report of mother's highest grade completed. Mother's education is more satisfactory for this purpose than father's because: (1) as we point out in Chapter 5, there is reason to believe that the mother is more influential in determining the linguistic skills and educational aspirations of her children than is the father; (2) in the case of broken homes, the mother is more likely to have custody of the children; (3) the subject is more likely to know the mother's than the father's education; (4) the subject is more likely to answer the question. Classification by mother's education is based on the 1979 interview question, "What was the highest grade or year of regular school that your mother ever completed?" The breakdown of responses to this question is shown in Table 2.4-7. For purposes of our analysis we have collapsed Mother's Highest Grade Completed into 0-8, 9-11, 12, 13-15, and 16 or more years.

TABLE 2.4-7  
PROFILE SUBJECTS BY MOTHER'S HIGHEST GRADE COMPLETED

Mother's Highest Grade Completed	Number of Subjects
0	117
1	22
2	87
3	184
4	162
5	199
6	415
7	265
8	798
9	690
10	987
11	1,083
12	4,602
13	361
14	536
15	173
16	630
17 plus	224
No information	374
TOTAL	11,914

## 2.5 Population Composition

We can regard the NLS sample as a sort of mini-census of the United States population between the ages of 15 to 22 in the spring of 1980. As such, the proportions of subjects in certain subclasses of the seven-way classification, suitably weighted to account for allocation sampling, should approximate to those of the Current Population Survey (CPS) around that time. In particular, figures are available in the CPS that enable us to check the important sociocultural groups by age by highest grade completed distribution of cases.

The weights for reconstituting a probability sample were calculated to match the proportions of Whites, Blacks and Hispanics, male and female, in the relevant age cohort as given in the CPS. This weighting takes into account any differential loss from these groups in the sample at screening and interviewing. [Details are given in Frankel and McWilliams (1981).]

The data presented in the tables in this section are based on the weighted 1979 NLS interview sample and not on the slightly less complete sample of subjects who actually took the tests in 1980.

Sociocultural Group by Economic Status and Region at Age 14. We see in Table 2.5-1 that, notwithstanding the internal migration following World War II, the number of Blacks relative to non-Blacks is still substantially greater in the Southeast than elsewhere in the country. And, by the same token, Blacks are relatively more numerous in the Northeast than in the Midwest or West. (Refer to Table 2.4-6 for the states included in these regions.)

We see also that the influx of Cubans into Florida, while locally concentrated, has not produced any appreciably greater proportion of Hispanics in the Southeast than in the Midwest. A somewhat larger proportion of Hispanics is found in the Northeast, but their relative numbers continue to be much greater in the Western states.

TABLE 2.5-1

PERCENT<sup>a</sup> OF NLS YEAR 1 SAMPLE IN EACH SOCIOCULTURAL GROUP AND ECONOMIC STATUS, BY REGION OF RESIDENCE AT AGE 14

Sociocultural/ Economic Group	Region of Residence at Age 14				All Regions
	Northeast	Southeast	Midwest	West	
Poor White	7.6	8.6	8.4	6.7	7.9
Nonpoor White	73.5	63.3	81.6	68.0	72.4
Poor Black	4.5	12.1	2.8	3.1	5.5
Nonpoor Black	8.8	14.1	5.5	5.8	8.4
Poor Hispanic	2.3	0.3	0.3	4.9	1.7
Nonpoor Hispanic	3.3	1.6	1.3	11.6	4.1
TOTAL	100.0	100.0	99.9	100.1	100.0
Percent in region	23.8	23.8	30.9	21.5	100.0

<sup>a</sup>Weighted.

The proportions of people classified as in poverty varies among the regions and the sociocultural groups. We can calculate from Table 2.5-1 that among Whites, 9.8 percent are in poverty; among Blacks, 39.6; and among Hispanics, 29.3 percent.

Highest Grade Completed by Region. Table 2.5-2 shows us that education level (as highest grade completed) in this sample is not as variable among these regions as one might suppose. The Northeast continues to send the greater proportion of its young people to college, especially when compared to the Southeast. The Southeast also has relatively more subjects with less than 9 years of education, the percentage in that group being nearly twice the percentage in the 0 through 8 group in the Northeast and Midwest. Other than these differences, regional variation in highest grade completed is not especially striking. For example, the proportions of subjects with 12 or more years of schooling varies only a little more than 10 percent. In terms of numbers of pupils completing high school, all regions are doing about equally well.

TABLE 2.5-2

PERCENT<sup>a</sup> OF NLS YEAR 1 SAMPLE COMPLETING EACH GRADE  
CATEGORY BY REGION OF RESIDENCE AT AGE 14

Highest Grade Completed	Region of Residence at Age 14				All Regions <sup>b</sup>
	Northeast	Southeast	Midwest	West	
0-8	5.2	8.6	4.8	7.0	6.3
9-11	41.3	48.1	45.1	46.9	45.3
12	30.4	27.9	33.1	27.9	30.1
13 plus	23.1	15.3	17.0	18.2	18.3
TOTAL	100.0	99.9	100.0	100.0	100.0
Percent in Region	23.8	23.8	30.8	21.6	100.0

<sup>a</sup>Weighted.

<sup>b</sup>This marginal distribution differs slightly from those in Tables 2.5-3 and 2.5-10 because complete classification data were not available for some cases in the sample.

Highest Grade Completed by Age. It is important to understand that the overall proportions of subjects from the Profile sample at the various educational levels cannot be interpreted in Table 2.5-2 because of the subjects in the sample who are too young to have completed high school or entered college. More useful figures are shown in Table 2.5-3 where the educational levels for subjects in three age ranges are shown. As mentioned above, the actual basis for the age classification is date of birth: reading left to right the age categories represent subjects born in 1962 and later, in 1960-61, and in 1959 and earlier, corresponding roughly to the ages 15 through 17, 18 and 19, and 20 through 22. Table 2.5-3 shows clearly that those members of the sample aged 15 through 17 who have completed 8 or fewer grades are merely behind for their age and will eventually move at least to grade 9 and possibly beyond. This is apparent in the very small numbers of subjects in the lowest educational level among those 18 years of age and older.

Highest grade completed among the 18- and 19-year-olds and the 20- through 22-year-olds can also be obtained from the Current Population Reports

TABLE 2.5-3

PERCENT<sup>a</sup> OF NLS YEAR 1 SAMPLE COMPLETING EACH  
GRADE CATEGORY, BY AGE GROUP

Highest Grade Completed	Age Group			All Ages <sup>b</sup>
	15-17	18-19	20-22	
0-8	12.8	3.0	3.2	6.6
9-11	86.2	31.6	14.7	45.0
12	1.0	50.5	43.7	29.9
13 plus	0.0	14.9	38.4	18.5
TOTAL	100.0	100.0	100.0	100.0
Percent in Age Group	36.4	25.2	38.3	99.9

<sup>a</sup>Weighted.

<sup>b</sup>This marginal distribution differs slightly from those in Tables 2.5-2 and 2.5-10 because complete classification data were not available for some cases in the sample.

for 1978 and 1979. That these figures, shown in Table 2.5-4, agree within a few percentage points with the figures in Table 2.5-3 is an indication that the NLS probability sample closely mirrors the population it is intended to represent.

Highest Grade Completed by Sociocultural Group and Economic Status.

Comparable to the regional variation in highest grade completed is the variation between the sociocultural groups shown in Table 2.5-5. In terms of the proportions in each group with high school education or better, the percentages range from 52.0 in the nonpoor Whites to 24.2 in the poor Hispanics. In terms of the proportion with some college education, however, the special nature of the economic status classification mentioned above comes into play. The sample contains a number of White college students who are living independently and are considered heads of households. Their income as students is low enough for them to be classified as poor by the criteria used here, even though this may be a temporary poverty. The result, however, is

TABLE 2.5-4

PERCENT OF NONINSTITUTIONAL UNITED STATES POPULATION  
COMPLETING EACH GRADE CATEGORY BY AGE GROUP<sup>a</sup>

Highest Grade Completed	Age Groups	
	18-19	20-22
0-8	3.5	3.4
9-11	34.7	12.7
12	49.3	45.3
13 plus	12.5	38.6
TOTAL	100.0	100.0

<sup>a</sup>Figures based on the March 1979 Current Population Survey. See U.S. Bureau of the Census, Educational Attainment in the United States: March 1979 and 1978, Current Population Reports, Series P-20, No. 356, Table 2, p. 20, August 1980.

TABLE 2.5-5

PERCENT<sup>a</sup> OF NLS YEAR 1 SAMPLE COMPLETING EACH GRADE CATEGORY,  
BY SOCIOCULTURAL GROUP AND ECONOMIC STATUS

Highest Grade Completed	White		Black		Hispanic		All Groups	
	Poor	Non-Poor	Poor	Non-Poor	Poor	Non-Poor	Poor	Non-Poor
0-8	15.5	4.2	12.9	6.2	25.4	12.0	15.8	4.8
9-11	42.7	43.8	58.6	45.6	50.5	49.0	49.4	44.3
12	22.6	32.1	19.3	30.4	15.2	23.6	20.5	31.5
13 plus	19.3	19.9	9.2	17.8	9.0	15.4	14.4	19.4
TOTAL	100.0	100.0	100.0	100.0	100.1	100.0	100.1	100.0
Percent in Group <sup>b</sup>	7.9	72.0	5.5	8.3	1.9	4.4	15.3	84.7

<sup>a</sup>Weighted.

<sup>b</sup>This marginal distribution differs slightly from that in Table 2.5-9 because complete classification data were not available for some cases in the sample.



that the nonpoor Whites and poor Whites show about the same percentage of subjects completing one of more years of college.

The genuinely poor Whites are, of course, much more represented among those not completing more than 8 years of school. This group constitutes 15.5 percent of the poor Whites. Among the nonpoor Whites 4.2 percent fall in this lowest educational level, and most of these are behind for their age and will eventually reach at least the 9th grade.

In all, the distribution of educational levels for the nonpoor Blacks is somewhat higher than that for the nonpoor Hispanics and not very different from that of the nonpoor Whites. The poor Blacks, and to an even greater extent the poor Hispanics, have the least favorable distribution of education. These are quantitative comparisons that do not necessarily reflect quality of education. We will come back to this point when we discuss sociocultural group differences in responses to the vocational tests in Chapter 5.

Highest Grade Completed by Sociocultural Group by Age. The sociocultural differences in highest grade completed are shown in a more detailed cross-classification by age group in Table 2.5-6. Comparing Table 2.5-6 with Table 2.5-7, which contains corresponding percentages calculated from the Current Population Survey, we see a possible tendency to undersample low education subjects from the Hispanic population aged 20-22. The lowest education group is underrepresented by 5.3 percent and the highest overrepresented by 10.3 percent. In spite of this problem in sampling the low education level Hispanics, we do not expect our analysis to be much affected. As mentioned in Chapter 1, the group comparisons, based on the interaction plots in Chapter 7, assume arbitrary numbers of subjects in the subclasses, and are not influenced by differences between the sample and population proportions. They would be affected only if there were differential misrepresentation of high and low ability subjects within education levels, and any such effects are likely to be small.

TABLE 2.5-6

PERCENT<sup>a</sup> OF NLS YEAR 1 SAMPLE COMPLETING EACH GRADE  
CATEGORY BY SOCIOCULTURAL GROUP AND AGE GROUP

Highest Grade Completed	White			Black			Hispanic		
	15-17	18-19	20-22	15-17	18-19	20-22	15-17	18-19	20-22
0-8	10.8	2.5	2.4	18.9	2.8	3.3	22.9	9.9	13.1
9-11	88.2	28.5	12.3	79.7	42.3	25.5	75.8	47.7	23.0
12	1.0	53.1	45.0	1.5	43.1	40.7	1.2	34.4	32.7
13 plus	0.0	15.9	40.2	0.0	11.8	30.6	0.1	8.0	31.3
TOTAL	100.0	100.0	99.9	100.1	100.0	100.1	100.0	100.1	100.1

<sup>a</sup>Weighted.

TABLE 2.5-7

PERCENT OF NONINSTITUTIONAL UNITED STATES POPULATION COMPLETING EACH GRADE  
CATEGORY BY SOCIOCULTURAL GROUP AND AGE GROUP<sup>a</sup>

Highest Grade Completed	White			Black			Hispanic		
	18-19	20-22	18-19	20-22	18-19	20-22	18-19	20-22	20-22
0-8	3.3	3.4	4.9	3.9	10.5	18.4			
9-11	32.7	11.5	48.7	21.9	42.5	20.1			
12	51.4	45.4	35.5	45.4	38.5	40.2			
13 plus	12.7	39.8	10.8	28.7	8.5	21.0			
TOTAL	100.1	100.1	99.9	99.9	100.0	99.7			

<sup>a</sup>Figures based on the March 1979 Current Population Survey. See U.S. Bureau of the Census, Educational Attainment in the United States: March 1979 and 1978, Current Population Reports, Series P-20, No. 356, Table 2, p. 20, August 1980.

Highest Grade Completed by Sex. There is an interesting trend over educational levels in Table 2.5-8. Relatively more males are held back or drop out without completing more than 8 years of education, and relatively fewer males go on to one or more years of college. This effect is especially prominent among Blacks, where 17.6 percent of females reported one or more years of college, as opposed to 10.9 percent of Black males. Some reasons for the greater educational attainment of Black women are discussed in Chapter 7. In the other sociocultural groups, the sex effect on the distribution of highest grade completed is in a similar direction as in Blacks, but much weaker.

Mother's Education by Sociocultural Group and Economic Status. The other important two-way associations in the population classification involve mother's education. As might be expected, the variations in mother's education with respect to region and sociocultural group are similar to those of their children, except stronger because they reflect origins of the parents and the state of education a generation earlier. Differences among sociocultural groups shown in Table 2.5-9 are especially dramatic, the most striking figure being 68.7 percent of poor Hispanic mothers reported to have no more than 8 years of education.

This figure probably reflects the agrarian and working-class origins of many of the poor Hispanic mothers, especially those from Puerto Rico and Mexico, where at the time (circa 1960) it would not be unusual for such girls to have only a grade-school education. We will have more to say about these kinds of migrational effects in Chapter 5.

The data in Table 2.5-9 also show more clearly than elsewhere the connection between the classification of households by income level and characteristics of parents of the Profile subjects. The nonpoor Whites have a

TABLE 2.5-8  
PERCENT<sup>a</sup> OF NLS YEAR 1 SAMPLE COMPLETING EACH GRADE  
CATEGORY BY SOCIOCULTURAL GROUP AND SEX

Highest Grade Completed	White		Black		Hispanics		All Groups	
	Male	Female	Male	Female	Male	Female	Male	Female
0-8	6.3	4.6	11.2	6.9	17.2	14.9	7.6	5.6
9-11	44.6	42.8	51.6	48.9	50.0	48.9	45.9	44.1
12	30.5	31.9	26.3	26.0	20.6	21.4	29.3	30.5
13 plus	18.5	20.7	10.9	17.6	12.2	14.7	17.1	19.9
TOTAL	99.9	100.0	100.0	100.0	100.0	99.9	99.9	100.1
Percent in each	40.7	39.2	7.0	6.9	3.2	3.1	50.9	49.2

<sup>a</sup>Weighted.

TABLE 2.5-9  
PERCENT<sup>a</sup> OF NLS YEAR 1 SAMPLE WHOSE MOTHERS COMPLETED EACH  
GRADE CATEGORY, BY SOCIOCULTURAL GROUP AND ECONOMIC STATUS

Mother's Highest Grade Completed	White		Black		Hispanic		All Groups	
	Poor	Non-Poor	Poor	Non-Poor	Poor	Non-Poor	Poor	Non-poor
0-8	20.2	7.3	23.1	11.4	68.7	44.5	27.2	9.6
9-11	24.0	15.7	42.6	32.1	16.1	19.5	29.6	18.3
12	40.3	51.9	28.0	37.3	11.1	25.2	32.3	49.1
13-15	8.7	12.5	4.4	11.5	2.6	6.0	6.4	12.1
16 plus	6.8	11.6	2.0	7.8	1.4	4.9	4.4	10.9
TOTAL	100.0	100.0	100.1	100.1	99.9	100.1	99.9	100.0
Percent in Group <sup>b</sup>	7.8	72.5	5.3	8.2	1.9	4.3	15.0	85.0

<sup>a</sup>Weighted.

<sup>b</sup>This marginal distribution differs slightly from that in Table 2.5-5 because complete classification data were not available for some cases in the sample.

much more favorable distribution of mother's education than do poor Whites, whose distribution across levels of mother's education resembles that of the nonpoor Blacks.

Mother's Education by Offspring's Education. Not surprisingly, a mother's education has a strong association with that of her children. Table 2.5-10 shows the distribution of highest grade completed among the NLS sample in each of the five mother's education groups. Having a mother with some college education greatly increases the probability that the subject will have some college education. Inasmuch as only about 50 percent of the sample had reached college age when this information was obtained, the effect of mother's education on the probability of her children attending college is even stronger than Table 2.5-10 indicates. We shall see in Chapter 6 that a mother's education is also strongly associated with the test performance of her offspring.

TABLE 2.5-10  
PERCENT<sup>a</sup> OF NLS YEAR 1 SAMPLE COMPLETING EACH  
GRADE CATEGORY, BY GRADES COMPLETED BY MOTHER

Subject's Highest Grade Completed	Mother's Highest Grade Completed					All Grades <sup>b</sup>
	0-8	9-11	12	13-15	16 plus	
0-8	17.7	9.7	3.9	2.8	1.8	6.4
9-11	47.5	53.4	43.0	39.0	40.0	44.9
12	27.1	28.9	34.7	23.1	20.8	29.9
13 plus	7.8	8.1	18.4	35.0	37.4	18.8
TOTAL	100.1	100.1	100.0	99.9	100.0	100.0
Percent in each mother's grade category	12.2	20.1	46.5	11.2	10.0	100.0

<sup>a</sup>Weighted.

<sup>b</sup>This marginal distribution differs slightly from those in Tables 2.5-2 and 2.5-3 because complete classification data were not available for some cases in the sample.

## CHAPTER 3

### THE TESTS

To understand the results of the Profile study, one must be familiar with the tests that make up the ASVAB. These tests, like all such psychological measures, can be described from a number of points of view--history of development, scientific rationale, construction and use, content, psychometric characteristics, factorial structure, and correlations with other measures. Typically in a report of this kind, samples of the test forms would be included in an appendix so that the reader could view the item content first-hand. Because the version of the ASVAB administered in this study is still in active use, however, its items cannot be publicly circulated at the present time. In Section 3.4 we have therefore simulated items in each of the tests and, for eight of the ten tests, displayed these items on a scale indicating their relative difficulties for the subjects in the weighted sample.<sup>1</sup> For the remaining two tests, which cannot be thus scaled (see Section 3.5), we show the practice items that actually appear in the test booklet. These displays and other information presented in this chapter give some idea of the actual skills and knowledge measured by the tests. Sample items taken from earlier versions of the ASVAB can be found in Maier and Fuchs (1973).

#### 3.1. History of the ASVAB

The Armed Services Vocational Aptitude Battery consists of ten tests resembling in many respects those in the well-known General Aptitude Test Battery (GATB) or the Differential Aptitude Tests (DAT). The ASVAB is the latest of a long line of selection and classification instruments developed by

-----

<sup>1</sup>The estimated item thresholds for these displays are taken from Bock and Mislevy (1981).

the Armed Services beginning with the famous Army Alpha and Beta tests of World War I. By World War II, the Army Alpha and Beta had evolved into the Army General Classification Test (AGCT), which was made up of three subtests-- verbal reasoning, numerical ability and spatial reasoning. Recruits in the World War II mobilization were screened with this test to ensure a minimum level of reading and general learning ability.

As the need for detailed job classification became more evident following World War II, special ability tests including shop mechanics, automotive information, radio and electrical information were added to the Army and Navy classification batteries. The Army, Navy and Marine Corps used these batteries to assign recruits to the numerous military occupational specialties and training schools. The Services carried out many studies of the validity of the tests in predicting final course grades and instructor ratings in these schools. Correlations between the predictor tests and these criteria were consistently high enough to justify continued use of the tests. (Similar studies of the ASVAB are discussed in Section 3.3.)

Prompted by the Selective Service Act of 1948, Army psychologists developed on the model of the AGCT a selection test called the Armed Forces Qualification Test (AFQT). When introduced operationally in 1950, the AFQT consisted of the arithmetic reasoning, vocabulary, and spatial relations scales from the AGCT, but later the Navy's tool usage test was added to supply more practical content. The AFQT remained the initial selection instrument for all the Services through the Korean and Vietnam conflicts and up to 1973. In that year, the Services discontinued use of a common AFQT, and from 1973 through 1975, each Service estimated AFQT scores from its own classification test battery.

During the time when each Service had its own classification battery, the Air Force began using shortened versions of its tests in a voluntary testing program for high school students. High school seniors were given the opportunity



to take the test to find out whether they would qualify for the Air Force and, also, to learn something of their prospects for success in various vocations. This testing program was sufficiently popular to be broadened into a joint effort of the Services. The high school students were tested on a form of the ASVAB similar to that used in the Profile study.

In a further move to unify and reduce the cost of testing (by eliminating the administration of separate selection and classification tests to the same recruits), the Defense Department adopted a single battery, the present ASVAB, and implemented its use for service-wide selection and classification of recruits. Although some compromise in differential efficiency had to be accepted to accomplish this goal, satisfactory ASVAB Forms 6 and 7 were developed and put into operational use for selection and classification in 1976. Development of Forms 8, 9, and 10 was initiated at the same time.<sup>1</sup>

The tests used in the Profile study are those from the current operational ASVAB, Form 8A, which was introduced in October, 1980. It is entirely a paper-and-pencil instrument and contains 10 independently timed and scored subtests including, in order of administration:

<u>Name</u>	<u>Abbreviation</u>	<u>Number of Items</u>	<u>Time (Min.)</u>
1) General Science	GS	25	11
2) Arithmetic Reasoning	AR	30	36
3) Word Knowledge	WK	35	11
4) Paragraph Comprehension	PC	15	13
5) Numerical Operations	NO	50	3
6) Coding Speed	CS	84	7
7) Auto and Shop Information	AS	25	11
8) Mathematics Knowledge	MK	25	24
9) Mechanical Comprehension	MC	25	19
10) Electronics Information	EI	20	9

-----  
<sup>1</sup> Further details are available in a report entitled "History of the ASVAB," prepared by the ASVAB Working Group for the Principal Deputy Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics), March, 1980.

Numerous studies on various forms of the ASVAB have investigated their reliability, predictive validity, and comparability with standard batteries such as the GATB and DAT. In discussing the findings of these studies, we use some technical concepts of psychological measurement, the essentials of which, for the benefit of the general reader, we review in the following section.

### 3.2 The Rationale for Vocational Testing

Perhaps the best general reference on the subject of vocational testing remains Appraising Vocational Fitness by Means of Psychological Tests by Donald Super and John Crites (1962). (But see also Bray and Moses, 1972.) These authors point out that the use of objective tests to select a number of employees (or recruits) from a larger pool of applicants is justified by a considerable accumulation of empirical studies going back to the 1920s. These studies show, almost without exception, that objective test procedures of employee selection are more economical and superior to the traditional alternative--namely, discretionary judgments based on an interview, references and possibly school grades. The problem with the traditional procedures is their lack of uniformity, consistency, and objectivity, and the ever-present biases and prejudices to which arbitrary judgments are subject. The superiority of objective test methods is established by studies of the "validity" of the tests--that is, the power of the selection tests to predict one or more criteria of job success.

Ideally, a validity study consists of (1) administering the vocational selection tests to a random sample of subjects from the applicant pool, (2) using a test score or combination of the test scores as a prediction scale for a specific job, (3) putting the sample of applicants on the job and, after some period of training or work, obtaining a quantitative measure of job performance or success, and (4) examining the empirical relationship between the selection scale (predictor) and the performance or success scale (criterion).

Usually, the relationship between predictor and criterion is sensibly linear, and the predictive power can be expressed simply as the correlation coefficient between the predictor measure and the criterion measure. This correlation is called the "validity coefficient" of the selection instrument. If, as is usually the case, the predictor and criterion or some one-to-one transformations of them have an essentially bivariate normal distribution in the population, the practical value of the scale for selecting employees can be expressed by calculating the expected proportion of employees who will exceed a certain criterion score if they are above a certain selection score. Under the stated distributional assumptions, this proportion depends on the validity coefficient and the selection ratio, that is, the proportion of applicants who must be taken on to fill the available jobs. If the selection ratio is unfavorable to the employer in that a large proportion of applicants must be hired, then the validity coefficient must be high in order for the selection scale to be useful. If the selection ratio is favorable and only a small proportion of the applicant pool is to be hired, then even a low validity coefficient will substantially improve the probability of selecting a successful employee.

These relationships are illustrated by Figure 3.2-1. The scale of measurement of the prediction (on the abscissa) and of the criterion (on the ordinate) is assumed to be in standard units (mean=0; standard deviation=1). The percentage of subjects above any specified criterion score or selection score can be set by choosing the normal deviate corresponding to that percentile point of the normal distribution. Then by use of the table of the bivariate normal distribution (National Bureau of Standards, 1959)--or in the case of Figure 3.2-1, by a computing approximation--these deviates and the correlation coefficient will determine the expected proportion of employees who will fall above (or below) the criterion score when their selection score

# Correlation

$\rho = .2$

$\rho = .4$

$\rho = .6$

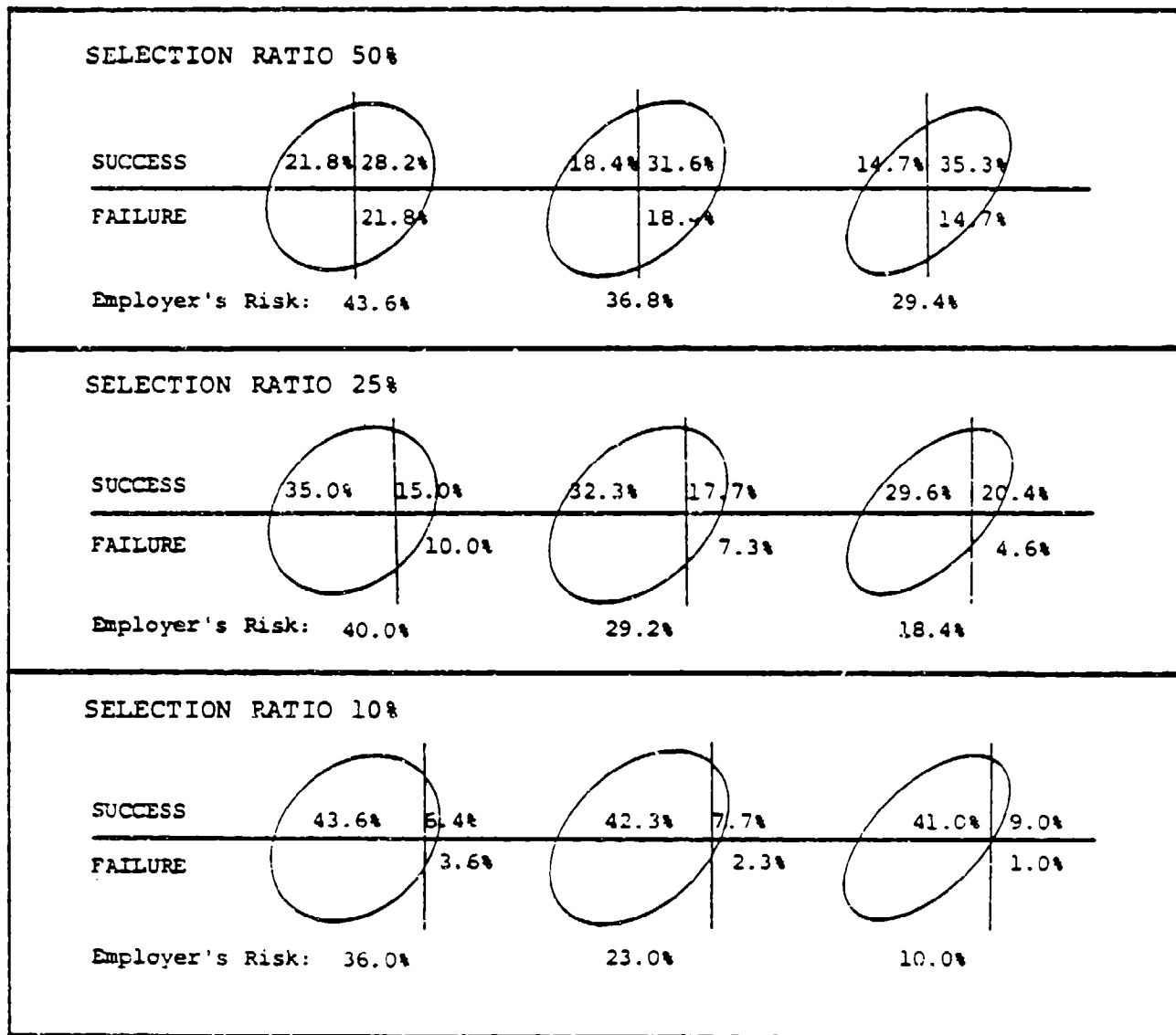


Figure 3.2-1 Employers' risks (percent of job failures among hired applicants) for various predictor validities ( $\rho = .2, .4, \text{ and } .6$ ) and selection ratios (percent of applicants hired)

is below the cutting point.<sup>1</sup> Of the four groups of applicants so assigned, two represent successful predictions, i.e., those simultaneously below both the selection and the criterion cutting scores or simultaneously above both the selection and the criterion cutting scores. The other two groups represent two types of errors of misclassification: Type 1 errors represent those persons who exceed the selection score but fail to surpass the criterion score; Type 2 errors represent those who fall below the selection score but above the criterion score. Obviously, Type 1 errors are more the concern of the employer because they represent losses due to hiring an unproductive worker. Type 2 errors are more the concern of the applicant who, because of misclassification, is erroneously excluded from employment when in fact he or she is capable of the work. We see in Figure 3.2-1 that, as the selection ratio becomes increasingly favorable to the employer, the Type 1 error is decreased at the expense of the Type 2 error. In other words, if the employer has only a small fraction of applicants to select, he can be fairly sure of selecting the right ones even with a selection test of modest validity. He will, of course, have passed over many applicants who may have done just as well, but since they cannot all be hired, the only rational course for the employer to take is to select employees from among those applicants who have the least risk of failure.

In specific situations the monetary benefits attributable to selection testing can be calculated. Maier and Fuchs (1973) for example, report an estimate placing the savings due to reduced training failure among Armed Services enlistees at 330 million dollars annually.

-----

<sup>1</sup>These proportions are also given in Taylor and Russell (1939).

The validation studies required to assess the error rates of a selection test are, perhaps not surprisingly, difficult to carry out in their ideal form. The greatest impediment in most cases is the impracticality of putting the entire sample of applicants on the job and observing who succeeds and who does not. Often the selection test is used to screen the validation sample so that only those applicants above a certain score level on the predictor are hired. This makes it impossible to know with certainty the true distribution of selection and criterion scores in the population or to estimate directly the validity coefficient. There exist, however, certain statistical approximations based on the theory of truncated normal distributions that can be used to recover estimates for the full distribution from the censored sample. In particular, a validity coefficient, "corrected for restriction in range," can be calculated as an approximation of the true coefficient for an unselected sample. These approximations are reasonably satisfactory, but they place greater demands on the distribution assumptions than would an analysis of complete data--especially so when the initial selection is severe and the adjustments in the validity coefficient are large.

There is occasionally an opportunity to examine the validity of selection tests in an unselected sample. An exemplary instance is the field validation study, carried out in 1953-54 by Colonel S.E. Jacobs (Retired), of the Examen Calificacion de Fuerzas Armadas (ECFA)--a Spanish language version of the AFQT. This test, along with a measure of English language knowledge and a non-language test of reasoning ability, was administered to a complete age cohort of male Puerto Ricans. Without regard to the selection scores, and excluding men only for medical, moral, or psychiatric reasons or legal deferment, the entire cohort was inducted and constituted as a special class in the training center at Tortugero. To prevent the wide range of aptitude among

these recruits from becoming too apparent and having an adverse effect on training, these inductees were assigned in a balanced manner with respect to ECFA scores to designated companies, platoons and squads by the investigating team and not by the training camp cadre.

Although some of these subjects, primarily from the low range of ECFA scores, could not adjust to basic training and had to be discharged, 528 completed sixteen weeks of basic training, including English language instruction. At the end of training, these men were administered a soldier performance test including demonstrations of physical proficiency as well as knowledge of equipment and procedures. In addition, each of the platoon sergeants, all experienced Puerto Rican soldiers, rated the troops in his platoon on a specially designed form. The trainees were also tested on their ability to comprehend and speak English and were rated on this criterion by their platoon sergeants.

Correlations between the pre-training and post-training measures from the study have been reported by Shenkel, et al. (1957) and also by Bock (1975). Table 3.2-1 shows the test intercorrelations calculated from a one-third random sample of the 528 cases as presented by Bock (1975). In comparison to other validation studies, this study produced very high relationships between the selection and criterion measures. The ECFA predicted soldier performance with a correlation of .71, and the English language pretest predicted post-test performance with a correlation of .93. Typical of subjective ratings as opposed to objective criteria, the corresponding correlations with the platoon sergeants' ratings of soldier performance in English language fluency were lower at .55 and .69, respectively.

This range of correlations, indicating a high degree of validity for predicting success in basic training, is comparable to or higher than the

TABLE 3.2-1  
CORRELATION OF BASIC TRAINING PREDICTORS AND CRITERIA  
FOR A SAMPLE OF PUERTO RICAN INDUCTEES  
(Bock, 1975)

Predictors	Criteria			
	Soldier Performance Test	English Fluency Test	Soldier Performance Ranking	English Fluency Ranking
Examen Calificacion de Fuerzas Armadas	.71	.79	.55	.71
Highest Grade Completed	.69	.76	.51	.71
Non-language Reasoning	.67	.69	.50	.62
English Fluency Test	.70	.93	.56	.69



validities for the ASVAB corrected for restriction of range discussed below. It gives us considerable assurance that the corrected figures from studies with restricted range are substantially accurate. Studies that do not report such corrections when the sample has, in fact, been selected on the predictor can seriously underestimate the validity coefficient.

### 3.3 Validity of the ASVAB

Typical validity coefficients for the ASVAB are seen in the study by Atwater and Abrams (1980). Aptitude composite scores were used to predict success in six Navy technical schools. Uncorrected and corrected validity coefficients for the schools were as follows:

	<u>Uncorrected</u>	<u>Corrected</u>
Air Traffic Controller	.45	.79
Aviation Support Equipment Technician	.19	.40
Basic Electricity and Electronics <sup>1</sup>	-.10	-.18
Signalman	.26	.48
Signalman 2	.37	.62
Basic Submarine	.27	.44

This is only one of many studies of the validity of the ASVAB tests and composites for predicting success in Service training schools. Vitola, Mullens and Croll (1973) reported corrected coefficients with Air Force training school grades ranging from .21 to .86 for airmen in forty-six different technical schools. Sims (1978) found for ASVAB Forms 6 and 7 corrected coefficients in the range .22 to .83 with a median of .60 for performance in 38 Marine Corps schools. As is typical of such studies, the correlations for skills that require assimilation and use of information are considerably higher than those for a mostly physical skill. Validity coefficients are also higher when the criterion measure is itself some kind of

-----

<sup>1</sup>The correlations are negative because the criterion measured time to complete training rather than final course grade.

paper-and-pencil instrument as opposed to a rating of actual performance. For example, Sims (1980) reports the following corrected correlations for Marine Corps training school grades: advanced auto mechanics, .74; electrical equipment repair, .73; field artillery control, .71. For more manual or physical occupations, however, the figures for ratings of performance are: tank crewmen, .22; correction specialists, .25; air crew service, .33; infantry training, .35; and military police, .37. Valentine (1977) found ASVAB validities of similar magnitude for final grades in forty-three Air Force technical schools.

Swanson (1979) compared ASVAB Form 6 and 7 scores of Navy enlisted men to final grades in 101 technical schools. The median correlation was .43 uncorrected and .73 corrected. For courses in which the recruits continued training until they reached a minimum competency level, the median correlation with time and training was -.21 uncorrected and -.36 corrected. But the various composites of the ASVAB do not differ very much in their predictive validity in the different schools, suggesting that it is mostly the general familiarity with the subject matter or general learning skill that is the main source of the battery's validity.

A limitation of many validity studies, both in the military and civilian sector, is that only success in training is evaluated and not performance on the job. An exception is the study by Larson and Aronson reported by Wilfong (1980), in which two samples of subjects who had voluntarily taken the ASVAB in high school were followed as they moved into the labor market. The first group consisted of 1,544 11th and 12th grade pupils completing the ASVAB Form 5 (high school version) in the fall of 1976; the second group consisted of 4,657 11th and 12th grade pupils taking the tests during the 1977-78 school year. To obtain criteria of job performance for these individuals, the investigators mailed questionnaires to those pupils who

had indicated during testing that their post-graduation plans included work, military service, or undecided. Among the questions was a request for the place of employment and the name and address of the immediate supervisor. These immediate supervisors were then sent questionnaires asking for ratings of on-the-job performance for the individuals they supervised. Based on questionnaires returned by the supervisors, multiple correlations between certain of the high school ASVAB tests and composites were calculated with the performance rating as criterion. The final sample size was not extremely large, and some "shrinkage" of these correlations was expected when the regression weights were applied to the ratings returned for the second group of subjects. Results for this cross-validation sample are shown in Table 3.3-1. These correlations have not been corrected for restriction of range due to selection because it was not possible in this study to know how selection was operating.

Considering the relatively crude criterion measure, the typical reluctance of supervisors to report poor performance and the absence of a correction for restriction of range, the validities of the ASVAB tests for success on the job are reasonably good. They must be viewed only as lower bounds on the actual validities, which probably are substantially higher. It seems safe to conclude that the ASVAB, although somewhat more abbreviated than a standard aptitude battery such as the DAT or GATB, is roughly comparable to those batteries in its power to predict job performance. Without going into more detailed consideration of how the test functions and what behaviors it draws upon in making these predictions, we have some assurance of its practical value for selecting recruits or employees with the best chance of being productive and successful.

TABLE 3.5-1  
SINGLE-SEX VALIDITIES OF ASVAB COMPOSITE SCORES AS PREDICTORS OF COMBINED PERFORMANCE CRITERION  
FOR JOB GROUPS DEFINED DURING CROSS-VALIDATION

(From Larson and Arenson, 1979)

Job Group	N	Verbal VE	Analytic/ Quantitative AQ	Clerical CL	Mechanical ME	Trade/ Technical TT	Academic Ability AA
Service Station Attendant (M)	63	.18	.18	.35**	.19	.09	.15
Driver (M)	133	.22**	.28***	.39***	.24**	.36***	.22**
Data Processing Worker (F)	69	.11	.21*	.10	.05	.10	.16
Guard (M)	43	.05	.17	-.18	.09	.28*	.04
Store Manager (M)	53	-.04	.14	.29**	.07	-.08	.16
Custodian (M)	50	.39**	.35*	.23	.37**	.40**	.37**
Child Care (F)	56	.14	.12	.27*	.22	.15	.14
Logger (M)	49	.34**	.35**	.32*	.36**	.30**	.28*
Sewing Machine Operator (F)	59	.11	.07	.02	.12	.17	.14
Auto Mechanic (M)	142	.16*	.27***	.19*	.15*	.16*	.20**
Plant Maintenance (M)	62	.26*	.28*	.31**	.30**	.27*	.23*
Electronic Technician (M)	92	.19*	.26**	.04	.15	.02	.17

\* Coefficient significant at .05 level

\*\* Coefficient significant at .01 level

\*\*\*Coefficient significant at .001 level

(M) Male Respondents Only

(F) Female Respondents Only

The research cited up to this point seems adequate to confirm the important role that differences in scores on the ASVAB will have on the prospects of young people for success in job training or job performance in either military or civilian life. We have not yet marshalled enough evidence to argue that the ASVAB is measuring the specific skills that are responsible for better job success among the higher scorers. All we know is that, directly or indirectly, the test is able to identify those who will so perform. Possibly the tests are just signaling the success of those persons who have a commitment to responsibility and hard work rather than some special abilities required for job success measured by these tests. We will discuss these questions in more detail in Chapters 5, 6 and 7.

#### 3.4 The Problem of Differential Predictive Validity

If a vocational test is used to select among persons with highly heterogeneous backgrounds and experience, it may happen that the relationship between the test score and a measure of job performance is different among distinguishable groups of such persons. The test is then said to have "differential predictive validity" such that different selection rules must be formulated for each group. An extreme example would be a population consisting of more than one language group, such as that of Belgium or Switzerland, where a test in the applicant's preferred language might have a different relationship to performance than a test in a second language. In such populations it may be necessary to construct and validate tests in more than one language and to allow the applicant to choose the language in which he or she will be tested. This is the policy, for example, in the Belgium armed services.

Another solution that has been attempted in these situations is to develop a non-language test that is valid for the criterion in question. A

number of tests purporting to measure general reasoning ability independent of language have been devised. Best known are the Raven Progressive Matrices and the Thorndike-Lorge Non-Language Tests. Regrettably, these tests, when applied to groups with differing linguistic and cultural backgrounds, typically show differential predictive validities similar to those of their language-based counterparts. It appears that the figural relationships upon which such tests depend are not equally comprehensible in groups with different exposures to such figures and relationships.

Less extreme but still consequential effects on predictive validity may also arise from experiential variation among persons who are nominally within the same language population. In a population as diverse as that of the United States, heterogeneity affecting predictive validity might exist at many levels. Whether it exists and in what degree is an empirical question that can be investigated at least for the larger subpopulations for which the necessary data can be obtained. In the U.S., studies of this type have been carried out for the major sociocultural groups--Whites, Blacks and Hispanics--and for the two sexes.

A differential predictive validity study requires a sample of applicants to be tested at time of selection and measured at a later time with respect to some criterion of job performance. On the basis of these data, the regression line showing the mean of the criterion scores for each value of the predicting test score is calculated by the method of least squares. In most applications it is satisfactory to assume a straight-line relationship in the criterion and predictor scores, in which case we may fit the regression lines by estimating the intercept and the slope in each of the groups in question. If the deviations of the criterion scores from the regression line within each group are normally distributed and have the same variance in all of the

groups, analysis of variance methods are available for testing the hypothesis that the regression lines are homogeneous from one group to another. The test has two component hypotheses: H1--equality of slopes between groups--and, given that H1 is accepted, H2--equality of intercepts between groups (see Bock, 1975, Chapter 7; Finn, 1978).

In the case of two groups, schematic regression lines are shown in Figure 3.4-1 representing the three possible outcomes of these tests of hypotheses--namely, accept H1 and H2, accept H1 and reject H2, and reject H1 and H2. Panel A of Figure 3.4-1 represents two homogeneous regressions (regression lines collinear--the length of each line represents 99 percent range of the criterion scores; on average, group 1 in this panel has lower scores on the predictor tests than does group 2, but it also has correspondingly lower average scores on the criterion). For predictive purposes, the tests are functioning in the same manner in both groups, the same regression line can be assumed in each group, and group membership of the applicant can be ignored during selection.

In panel B, the extent to which criterion scores depend upon predictor scores is the same for each group, but the level of performance is uniformly higher in group 2 (i.e., the regression lines are parallel but not coincident). The condition represented in panel B would occur when job performance depends upon two factors, one related to the predictor tests and the other not so related.

It is easy to imagine how this might occur. Suppose, for example, that a test of general verbal ability was used to predict some measure of skill in playing basketball among Black and White boys in an urban high school. It is likely that, within these groups, those Blacks with higher general verbal ability use better strategies and tactics and thus score higher

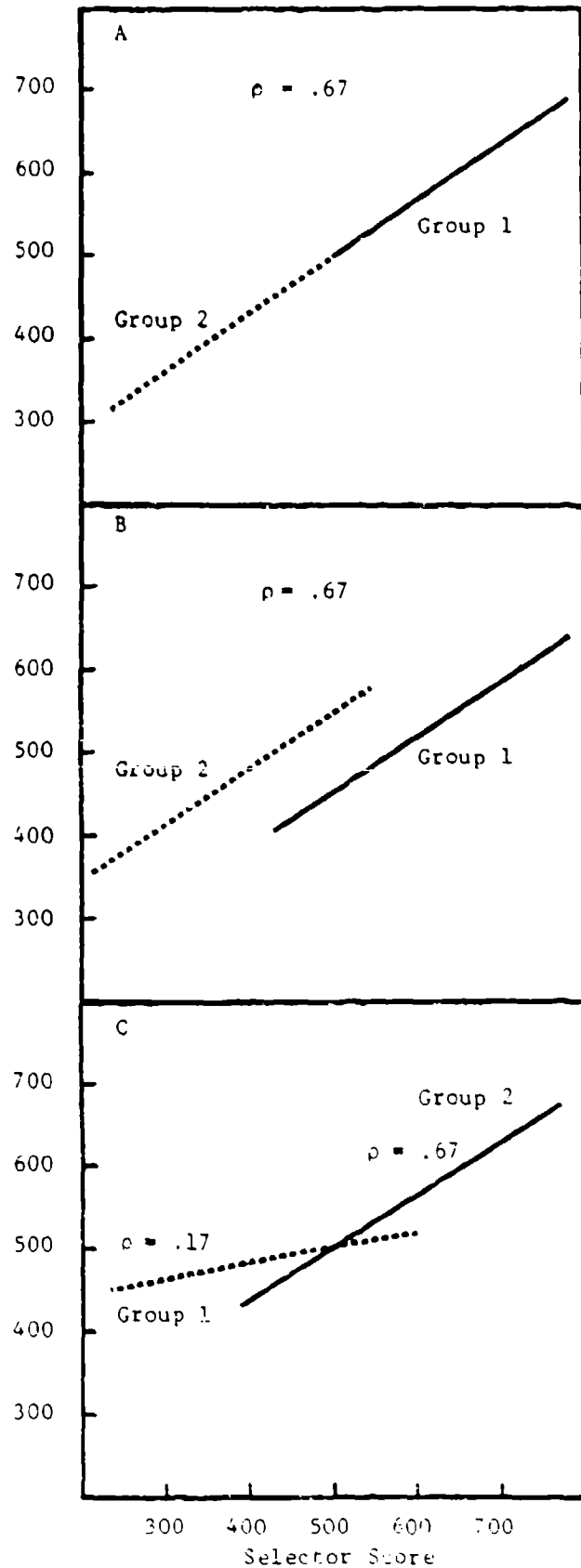


Figure 3.4-1 Schematic selector vs. criterion regression lines for two groups.  
A: No differential validity; B: Selection bias favoring group 1;  
C: Differential validity.



in play than those of lower general ability. At the same time, it may be the case that Blacks devote more time to playing basketball both in and out of school. Compared to Whites of the same general ability they could therefore be expected to score higher in play, other things being equal. If the amount of such extra practice is unrelated to general verbal ability, the upward displacement of the regression line for that group shown in panel B would result. For purposes of selection, a constant amount equal to the difference in intercepts divided by the slope should be added to the test scores of Blacks to obtain the best prediction of the criterion score. The Blacks and Whites thus selected would have different average general ability scores, but their performance in playing basketball would be equal.

We should also be alert to the possibility that deficiency in the design of the validation study can also produce the result shown in panel B. For example, selection may be influenced by some unknown factor that depends on group membership but is unrelated to predictor scores within groups. Suppose that selection is not done "blind" and that the person responsible for selection is permitted to use discretionary judgement in borderline cases. If these judgments were based on other information about the basketball playing ability of Blacks, so that in effect the Blacks were being screened more stringently than Whites, the ultimate performance of the Blacks would be higher at given selection scores. This would be especially likely to happen if there were many more Blacks than Whites in the school, and the person doing the selection was attempting to better balance the racial composition of the team. The importance of rigorous controls in carrying out a validation study, especially in the selection of applicants without regard to group membership, cannot be too much emphasized.

Finally, panel C depicts a situation in which the selection test is not related to the performance measure to the same extent in both groups. This result, which typically occurs when the groups differ in their average predictor scores, is symptomatic of technical shortcomings of the predictor tests. The tests may not have a uniform distribution of item difficulty, and may be less reliable at lower score ranges than at higher. This is especially a problem in multiple choice tests where the effects of guessing produce less reliable scores in the low range. The resulting non-uniform attenuation of the predictor-criteria correlation can lead to regression lines such as shown in panel C. The greater measurement error in Group 1 has reduced the slope of the regression line relative to that of Group 2 as shown in panel C. The effect of non-uniform reliability of the tests can, of course, be corrected by writing additional items at the required difficulty levels and repeating the validation study.

Another reason for the effect shown in panel C, and one more difficult to correct, is that the dimension measured by the test actually changes as the items increase or decrease in difficulty. It may be that the person constructing the test, in trying to devise more difficult items, added additional dimensions to the task that the items present. If these dimensions are relevant to criterion performance, the validity of the tests will increase as the item difficulty increases. In this case, scores in the lower ranges of the predictor will be less valid than those in the higher ranges, with the result that the regression lines will appear as in panel C. It is even possible that a test may be multi-dimensional throughout its range, and that some of these dimensions are predictors in one of the groups but not in the other. This would also tend to produce different slopes of the regression lines. These conditions could be corrected only by analysis of the item

content, possibly aided by item factor analyses in separate samples from the two groups. New items would have to be written, subjected to item analysis, and a test constructed that is equally valid, and thus has equal regression slopes, in groups at different score levels.

In principle, regression lines with unequal slopes can be used for selection by means of the Johnson-Neyman technique (see Johnson and Neyman, 1936) which assigns different selection rules at different levels of predictive score. The procedure is cumbersome, however, and is less reliable than single or parallel regression line models in comparable data. It has rarely been used in practical personnel selection.

In the interpretation of validity studies, it is of course not just the statistical significance of deviations from homogeneous regression that must be considered, but also the actual sizes of the deviations and the practical effect of possibly ignoring them in favor of a simpler selection rule. When the validation studies are done in very large samples, even minor departures from the assumption of homogeneous regression will be significant. But when the actual effects of the heterogeneity on the types of correct and incorrect classification are calculated by the methods discussed in Section 3.3, it might be found that the effects are too small to be of practical importance.

#### 3.4.1 Studies of Differential Validity of the ASVAB and Related Tests in Sociocultural Groups and the Two Sexes

The Armed Services have carried out two major studies of the possible differential validity of their classification tests. The first of these studies was carried out in 1978 by Milton Maier and Edmund Fuchs of the U.S. Army Research Institute for the Behavioral and Social Sciences. Subjects for this study were approximately 14,000 men in various military training programs

in 1964 and 1965. From these soldiers, samples of Black and White trainees were selected on the basis of their Army records for purposes of a differential validity study.

The second study was carried out by Lonnie Valentine, Jr., of the Air Force Human Resources Laboratory. Subjects for this study were men and women in the Air Force technical training schools between 1974 and 1977. These subjects were classified as White, Black, other, and as male and female for the purposes of the validity study.

Both these studies used as their main criterion of performance the final course grades in technical training schools operated by the corresponding services. In an earlier report, Maier and Fuchs (1973) discuss the reasons for choosing final training course grades as a measure of job performance in preference to evaluation of actual on-the-job behavior. They argue that because job duties vary so much from one assignment to another it is virtually impossible to obtain a general measure of job performance. A training course and its end-of-course tests sample most of the skill elements required on the job and thus provide a more representative index of how the soldier would perform in a wide variety of situations. In addition, performance in many jobs cannot be measured objectively and must be assessed by means of supervisors' ratings. There is abundant evidence in the literature of applied psychology, however, that supervisors' ratings are extremely unreliable and inconsistent. They are less dependable than the tests used to predict them. End-of-course grades in the military training schools, in contrast, are based on objective tests, usually with reasonably good reliability. Because results on these tests become part of each soldier's record along with his or her classification tests scores, it is relatively easy to gather together the data needed in a validity study.

Army Training Schools. Screening the records of some 25,000 soldiers entering military training programs in 1964 and 1965, Maier and Fuchs (1978) located approximately 14,000 men for whom racial identification and test scores for the Army Classification Battery (ACB) were available. The ACB consists of fifteen tests covering much the same ground as ASVAB but in greater detail. From the scores on these tests, the investigators calculated composite scores using formulas previously developed to predict job performance in the areas shown in Table 3.4-1. This table also shows the number of soldiers classified as White and Black for whom end-of-course grades were available in training programs pertaining to each group of military occupational specialities. (Data for other racial groups was too limited to be analyzed.)

The scores for both composite predictors and the course grades were standardized to a mean of 100 and standard deviation 20 in the total sample of 14,127 soldiers. Regressions for each MOS group were obtained in the following way: scores for each predictor were grouped into six successive equal intervals and the mean end-of-course grade was obtained for each. Straight trend lines were then fitted to each set of points with the length of the line indicating the range of scores in the data. Plots of the resulting trend lines, adapted from Maier and Fuchs (1978), are shown in Figure 3.4-2. Data for the field artillery groups were considered too limited to justify plotting.

Note that the slopes of the regression lines are in the range 0.5 to 0.75, indicating good validity of the composites for both White and Black soldiers. For two of the specialties, General Maintenance and Clerical, regression lines for Blacks and Whites are practically coincident and clearly indicate that the classification tests are functioning in the same way in both

TABLE 3.4-1

SAMPLE SIZES IN APTITUDE AREA VALIDATION AND RACIAL STUDIES

MOS Group	Sample Size			
	Validation	White	Black	White & Black
Combat Arms	1,609	581	131	712
Field Artillery	665	134	41	175
Electronics Repair	3,840	2,036	232	2,268
Operators and Food	1,516	1,166	179	1,345
Surveillance and Communications	2,137	1,243	148	1,391
Mechanical Maintenance	4,395	2,776	266	3,042
General Maintenance	1,139	688	76	764
Clerical	3,502	2,232	446	2,678
Skilled Technical	2,175	1,499	253	1,752
Total	20,978	12,355	1,772	14,127

Source: Milton H. Maier and Edmund F. Fuchs, Differential Validity of the Army Aptitude Areas for Predicting Army Job Training Performance of Blacks and Whites. Technical Paper 312. U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA, 1978, p.8.

groups. In three of the specialty groups, Electronics Repair, Operators and Food and Motor Maintenance, the regression line for Blacks has slightly lower slope than that for Whites, and it crosses the line for Whites near the mean for the Black group. Assuming that these differences in slope are due to reduced reliability or differences in content in the items at lower levels of difficulty, we might expect that if the test were improved in these respects the regression lines for Blacks and Whites would also become essentially coincident for these specialty groups.

Only in the case of Combat Arms, Surveillance and Communications, and the Skilled Technical specialties is there a tendency for the two regression lines to have slightly different slopes and to be sufficiently displaced from

LEGEND: ———— Validity study regression  
 - - - - - White sample trend line  
 - - - - - Black sample trend line

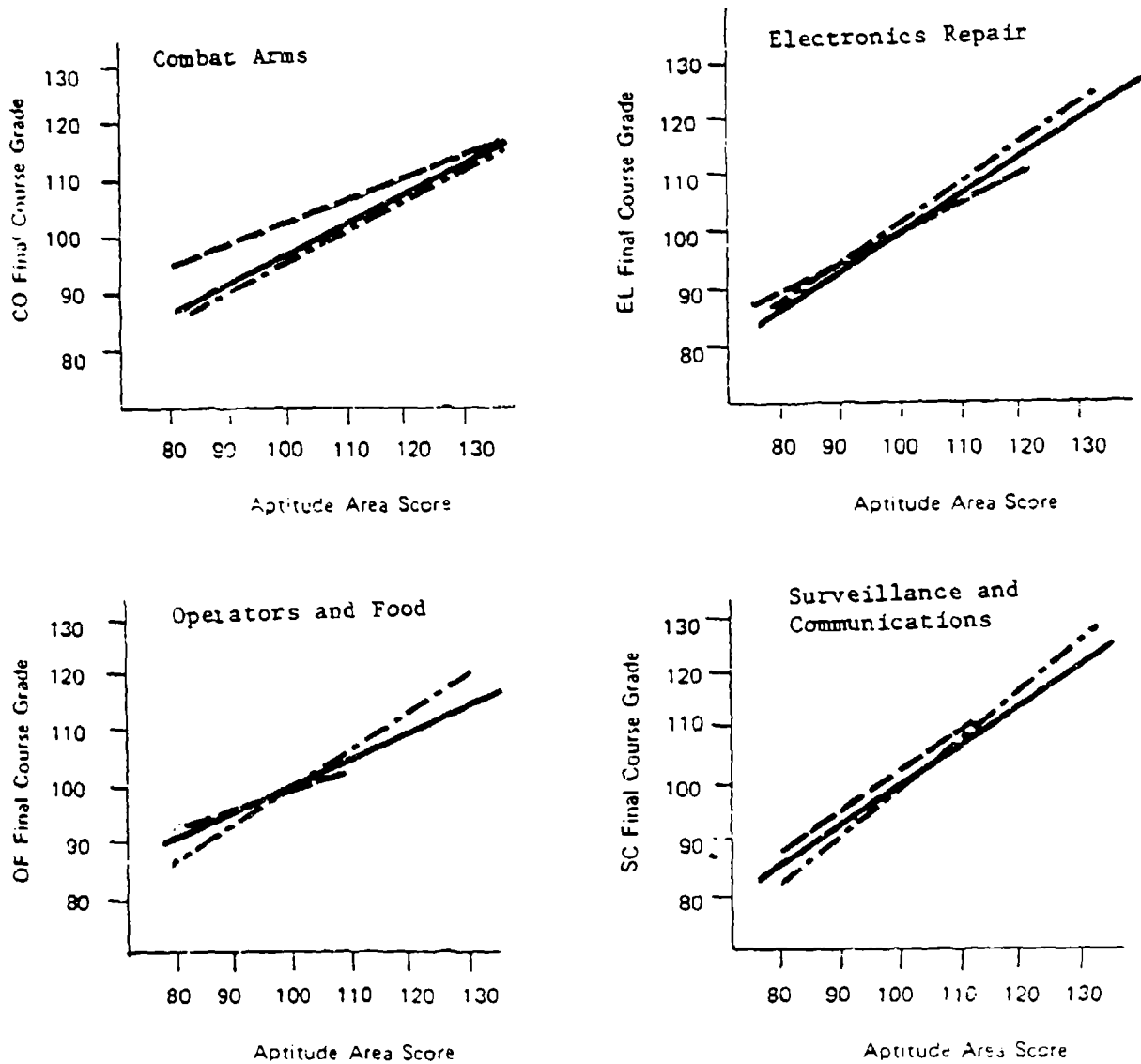


Figure 3.4-2 Comparative trends of course grades of Whites and Blacks for aptitude scores.

Source: Maier and Fuchs, 1978

LEGEND: ————— Validity study regression  
- - - - - White sample trend line  
- - - - - Black sample trend line

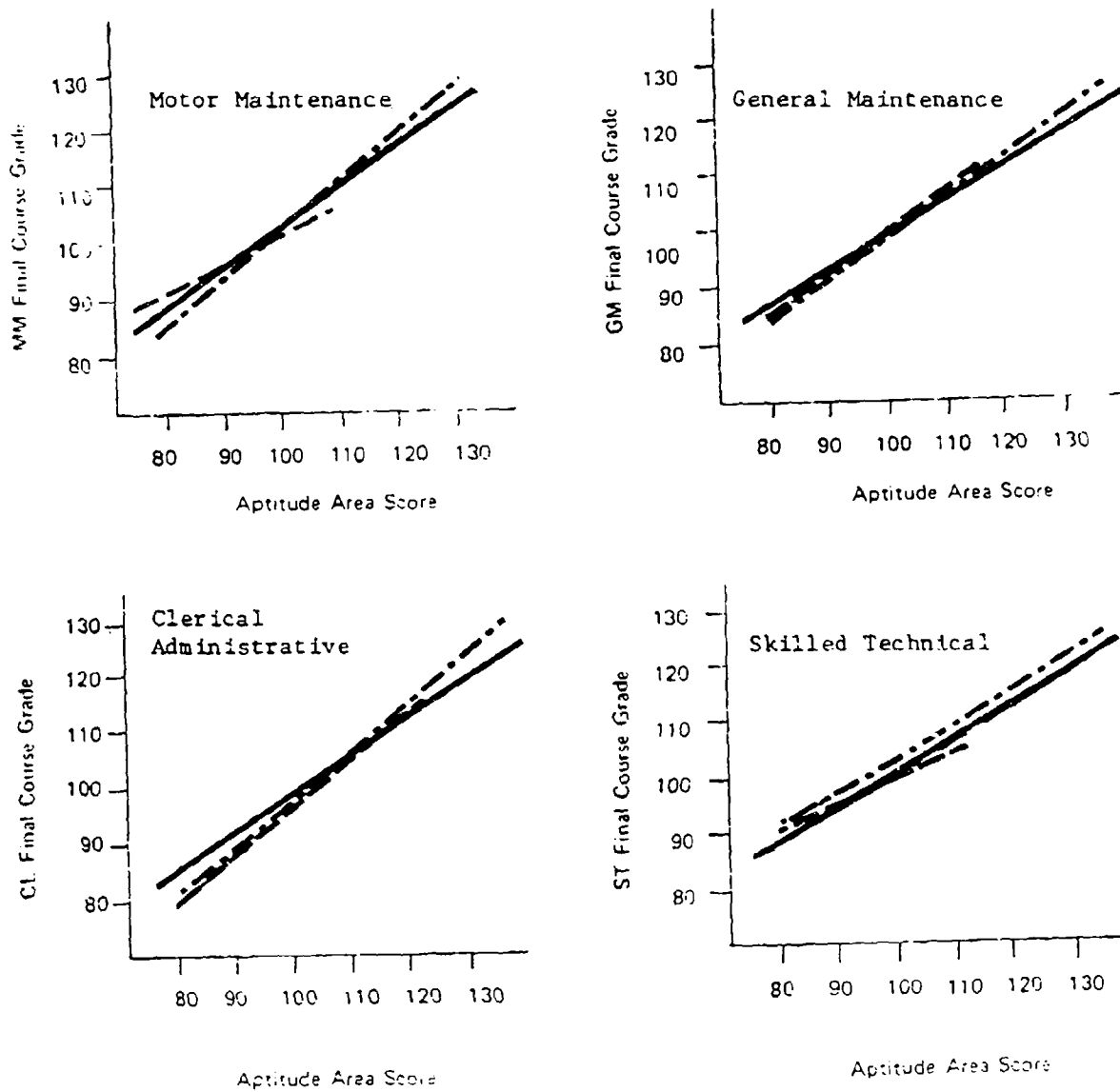


Figure 3.4-2 (cont'd).

Comparative trends of course grades of Whites and Blacks for aptitude scores.

Source:

Maier and Fuchs, 1978



one another as not to cross in the region of the score distributions. For Surveillance and Communications, the regression line for Blacks is above that for Whites, and for Skilled Technical below, but in each case the displacement and difference in slope is too small to be of any practical importance. In the Combat Arms specialties, which include Infantry, Armor, and Combat Engineer, the regression line for Blacks has a somewhat smaller slope than that for Whites and is clearly displaced upward. At the highest level of the Combat composite predictor, the mean final combat course grades for the groups are almost equal, while at the lowest level the mean for Blacks exceeds that for Whites by about 0.4 standard deviations.

This result suggests that, at least at lower levels of the predictor score, factors not accounted for by the test are acting to increase the final grades in the combat courses. Without knowing the content of the combat courses or how they were evaluated, it is difficult to know why in this area alone there is evidence of a non-negligible differential validity. Perhaps this is the area where the content of the course is furthest removed from the content of the predictor tests, thus giving more opportunity for other unrelated factors to affect final course grade. Certainly, the predicting composite includes such tests as Arithmetic Reasoning, Trade Information and Pattern Analysis, which may have limited relationship to combat course performance and on which most low-scoring Blacks may be at a special disadvantage. This is the only specialty area where this study might give some justification for separate selection rules for Blacks and Whites, other policy considerations permitting. In all of the other areas the regression lines for Blacks and Whites are so nearly the same that there would be no justification for distinguishing race during selection.

Air Force Training Schools. In his differential validity study of the ASVAB, Form 3, Valentine (1977) looked at forty-three Air Force technical training schools. For present purposes, we have extracted the results for the eight schools with larger sample sizes. Although Valentine gave results for other minority groups in some cases, we will restrict attention to differential validity for Black and White Air Force trainees, and male and female trainees. The names of the schools, numbers of subjects, and the ASVAB composite used for predicting final course grades are shown in Table 3.4-2. The same composites are used in more than one type of training school.

The tests of ASVAB 3 differ somewhat from those of ASVAB 8: Technical Knowledge replaces Science Information, Space Perception replaces Paragraph Comprehension, and Auto and Shop Information are two different tests. The other tests are basically the same in both forms.

Valentine examined the regressions of end-of-training-school grades on four predictor composites of ASVAB 3 test scores. He did not report means and standard deviations for these composites in the population of applicants, but these statistics can be calculated from the means, standard deviations and correlations of the ASVAB 3 tests in a large sample of applicants as reported by Seeley, Fischl and Hicks (1978). The cases in the latter sample were weighted so as to reproduce the AFQT distribution of the World War II mobilization population. The means and standard deviations of the composites computed from Seeley, Fischl and Hicks's Table 8 appear in Table 3.4-3 of the present section. The figures in Table 3.4-3 refer to composite scores calculated from the ASVAB 3 raw scores by the formulas shown in the table.

Results from Valentine's study for those schools with larger numbers of trainees are abstracted in Table 3.4-2. Figures for the smaller minority groups have been omitted. The means and standard deviations for composites

TABLE 3.4-2

SAMPLE SIZES, CORRECTED AND UNCORRECTED PREDICTOR-CRITERION  
CORRELATIONS, PREDICTOR COMPOSITE, AND FINAL GRADE RAW  
SCORE MEANS AND STANDARD DEVIATIONS FOR EIGHT AIR  
FORCE TRAINING SCHOOLS: REPORTED BY  
SOCIOCULTURAL GROUP AND SEX<sup>1</sup>

Service School	Sample Size	Correlation Cor.	Uncor.	Composite Mean	S.D.	Final Grades Mean	S.D.
<b>Communications- Electronic Systems</b>							
				(Elect)			
White	1,849	.89	.44	60.0	3.57	85.0	6.01
Black	181	.84	.34	58.2	3.42	81.8	5.34
Male	1,740	.89	.43	59.8	3.55	85.0	6.03
Female	343	.91	.47	58.3	3.56	84.0	5.92
<b>Avionics Systems</b>							
				(Elect)			
White	2,165	.81	.44	60.2	3.51	84.4	6.38
Black	244	.85	.35	59.0	3.41	81.1	6.44
Male	2,014	.83	.34	60.0	3.49	84.1	6.42
Female	450	.77	.27	58.5	3.39	84.0	6.59
<b>Aircraft Maintenance</b>							
				(Mech)			
White	4,559	.44	.34	70.4	14.2	84.4	7.78
Black	1,073	.19	.12	63.0	11.81	79.2	7.92
Male	4,468	.57	.38	73.2	11.35	84.0	7.99
Female	1,268	.16	.10	53.7	12.13	81.3	7.98
<b>Aircraft Engineer</b>							
				(Mech)			
White	1,356	.57	.40	72.0	12.2	85.3	7.41
Black	363	.57	.29	63.2	8.30	80.0	7.00
Male	1,431	.73	.49	71.7	10.02	84.4	7.74
Female	332	.00	.00	61.6	7.03	83.3	7.07

TABLE 3.4-2 (cont'd)

Service School	Sample Size	Correlation		Composite		Final Grades	
		Cor.	Uncor.	Mean	S.D.	Mean	S.D.
Transportation				(Adm)			
White	1,106	.34	.24	68.2	12.40	83.1	6.58
Black	400	.00	.00	66.5	10.50	79.1	5.94
Male	1,346	.28	.18	66.6	11.55	81.9	6.61
Female	200	.36	.21	76.9	10.12	83.3	7.07
Administration				(Adm)			
White	1,503	.38	.22	75.0	10.00	84.6	6.57
Black	1,078	.15	.08	70.8	9.62	81.7	6.34
Male	1,716	.34	.19	72.0	9.88	83.0	6.53
Female	921	.33	.19	75.4	10.05	84.2	6.69
Law Enforcement and Corrections				(Gen)			
White	1,078	.74	.39	53.8	6.42	83.6	5.90
Black	256	.64	.29	51.7	6.10	79.8	6.66
Male	900	.73	.38	53.9	6.55	83.6	6.19
Female	448	.75	.38	52.7	6.03	81.4	6.02
Medical				(Gen)			
White	1,385	.72	.37	50.0	5.47	82.4	6.98
Black	470	.66	.26	55.6	5.11	76.3	6.97
Male	1,283	.81	.42	57.6	5.56	80.9	7.49
Female	620	.70	.29	57.0	5.23	81.0	7.48

<sup>1</sup>From Valentine, L.D., Prediction of Air Force Technical Training success from ASVAB (Form 3) and educational background. Report No. AFHRL-TR-77-18. Brooks Air Force Base, Texas, 1977.

TABLE 3.4-3

ASVAB-3 PREDICTOR COMPOSITES: UNSELECTED RAW SCORE POPULATION  
MEANS AND STANDARD DEVIATIONS<sup>1</sup>

Composite	Composition <sup>2</sup>	Mean	S.D.
Electronics	(Elect) AR + SP + EI	41.3	14.44
Mechanical	(Mech) TK + MC + SI + AI	58.0	19.31
Administrative	(Adm) WK + CS	55.1	18.27
General	(Gen) 2 + WK + AR	41.4	16.87

<sup>1</sup>From Seeley, L.C., Fischl, M.A., and Hicks, J.M. (1978). Development of the ASVAB, Forms 2 and 3. Technical Paper 289, Army Research Institute, Alexandria, Virginia.

<sup>2</sup>AR = Arithmetic Reasoning

SP = Space Perception

EI = Electronics Information

TK = Technical Knowledge

MC = Mechanical Comprehension

SI = Shop Information

AI = Auto Information

WK = Word Knowledge

CS = Coding Speed

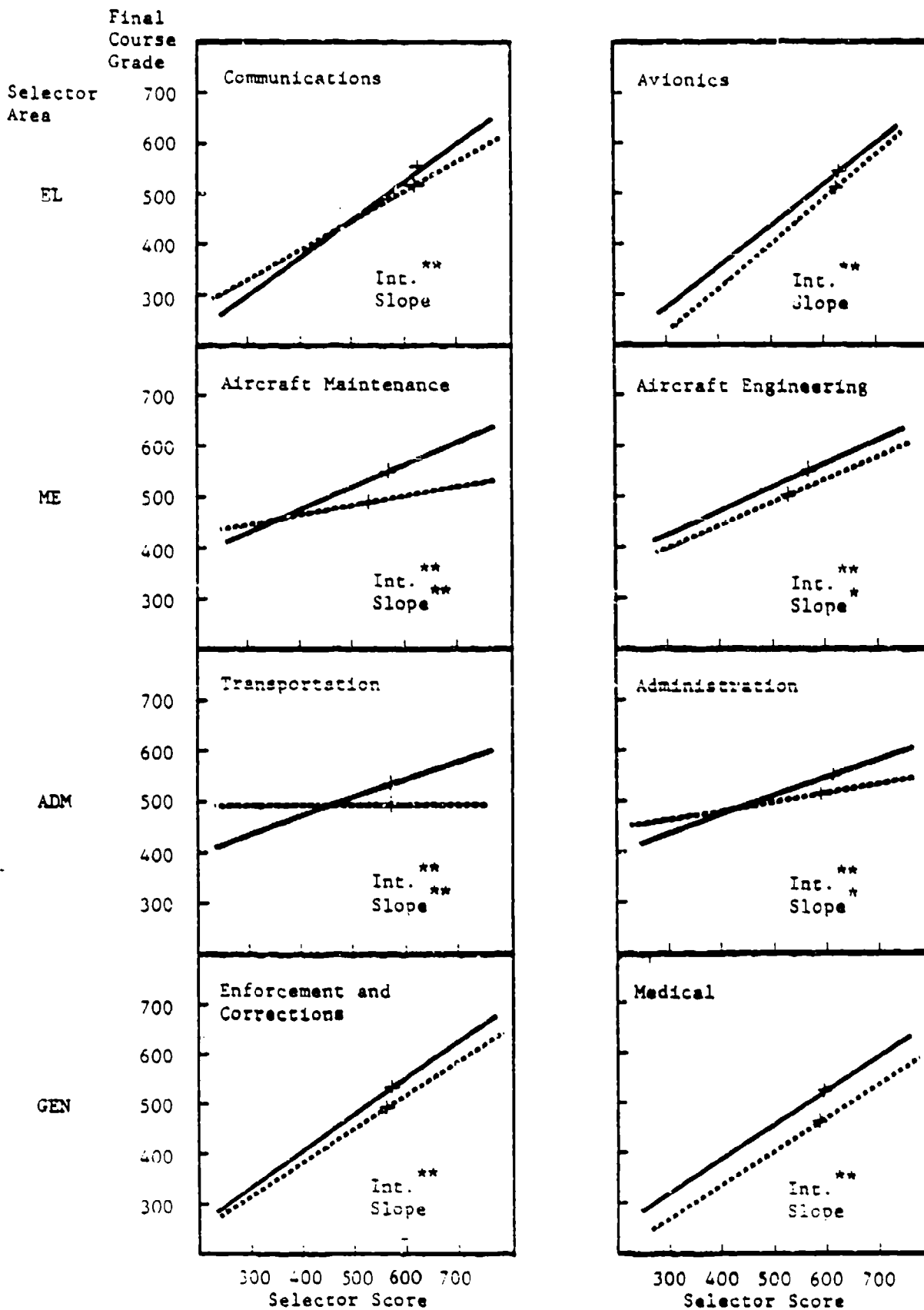


Figure 3.4-3 Regressions for White and Black trainees of Final Course Grade on selector composites for Air Force Training Schools. \*Slope or intercept significantly different at .05 level; \*\*at .01 level. Solid line=White; broken line=Black.

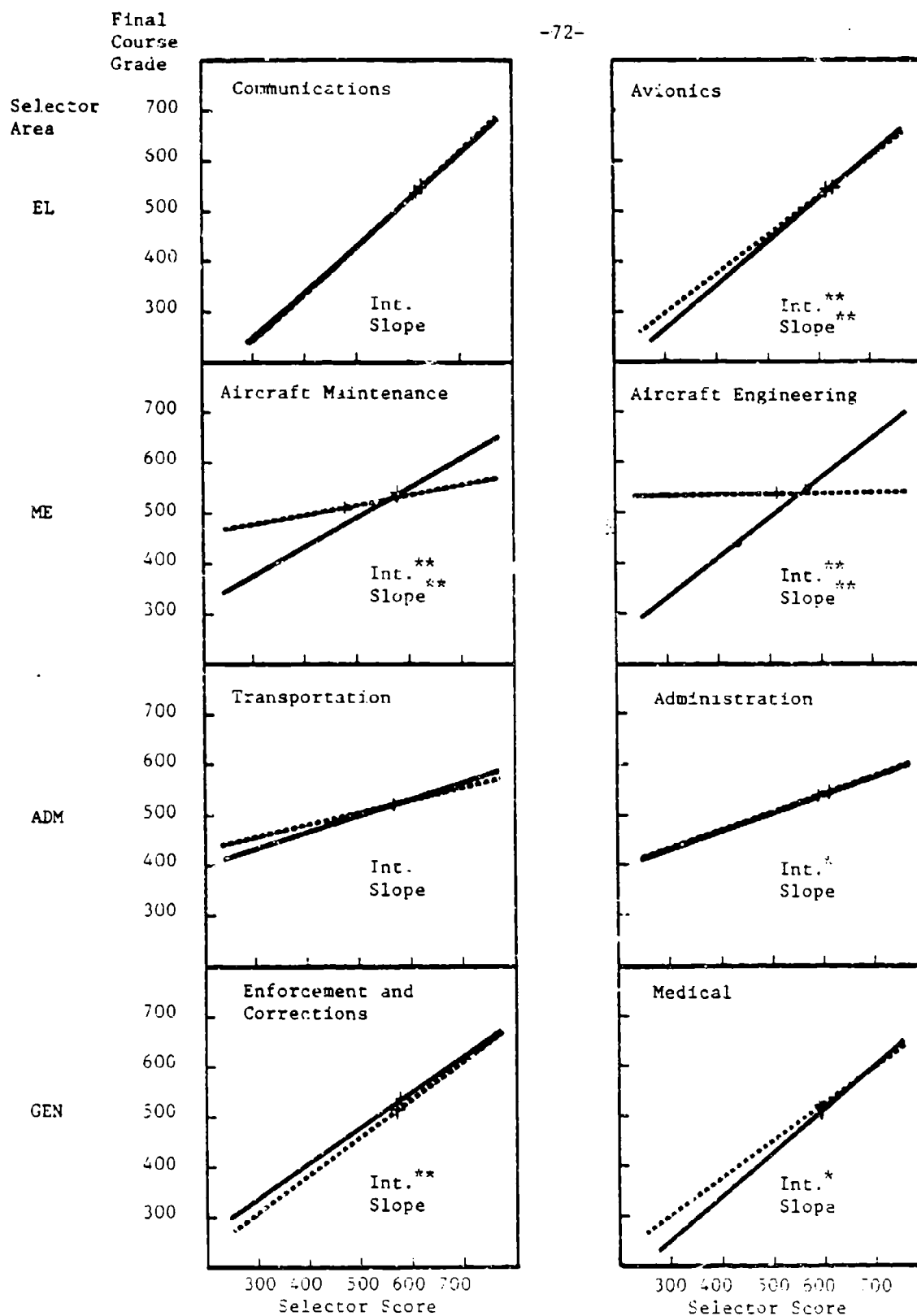


Figure 3.4-4 Regressions for male and female trainees of Final Course Grade on selector composites for Air Force Training Schools. Variance of selector corrected for restriction of range. \*Slope or intercept significantly different at .05 level; \*\*at .01 level. Solid line=Male; broken line=Female.

reported in percentiles by Valentine have been converted to raw score units for comparability with Table 3.4-3. In Table 3.4-2 correlations between each of the four composites--Electronic, Mechanical, Administrative, and General--and the final school grades of trainees selected by these composites are shown both in the uncorrected form given in Valentine's Table 2 and after correction for restriction of range. Comparing the means for the selected groups with the means and standard deviations in Table 3.4-3, we see that selection was quite stringent for all these schools, and especially so for the popular Avionics specialty, where the mean of the selected group is nearly one-and-a-half standard deviations above the mean of the general population. It should also be noted that the mean composite score for Blacks is less than that for White males. This indicates that a larger proportion of Blacks are to be found near the cutting point between selection or rejection regions. As we shall see later in this section, this fact has implications for the regression of final course grades on the predictor composite.

Regression lines for the White and Black, Male and Female groups based on Table 3.3-2 are expressed in standardized scores in Figure 3.4-3 and 3.4-4 (mean=500, standard deviation=100). Also shown are the results of significance tests for the difference in intercept at the mean and difference in slope calculated by Valentine without correction for restriction in range. One asterisk means that the corresponding F-statistic exceeded the 0.05 point; two asterisks means it exceeded the 0.01 point. Because of the small sample sizes, rather small differences are detected in some cases.

The regression lines in Figure 3.4-3 and 3.4-4 show no evidence whatsoever of biased selection favoring Whites to Blacks or men to women. Indeed, the lines for Avionics Systems, Aircraft Engineering, Law Enforcement and Corrections, and Medical might be construed as evidence of bias in favor



of Blacks. A more likely interpretation, however, is that the selection of Blacks near the cutting point (mentioned above) tends to include a greater proportion of chance successes than among Whites, whose distribution is shifted toward higher levels on the selection composite (see the distributions in Section 4.2). Because the selection tests are not infallible, some examinees near the cutting point are classified as capable of passing the training course when in fact their chances of doing so are poor. These cases regress to lower values on the final course grade, and bring down the average of their group as seen in the shifting downward of regression lines for Blacks in Figure 3.4-3.

Both Figures 3.4-3 and 3.4-4 show another kind of differential validity affecting the Mechanical and Administrative composites. Unlike the Electronics and General composites, which show excellent and virtually equal validities in all groups, the validities of the Mechanical and Administrative composites are rather poor for Whites and in some cases almost nil for Blacks and females. The slopes of the regression lines indicate that the Mechanical composite has almost no predictive validity for Blacks in the Communications-Electronics Systems course, and none for females in either of the mechanical specialties. The Administrative composite has little or no predictive validity among Blacks for the Transportation or Administration course. Other results reported by Valentine indicate that the General composite would have served much better to predict success in these courses.

These limitations of the ASVAB tests, although not unfair to Blacks or females as groups, are undesirable in individual cases when the composites falsely accept unqualified applicants or falsely reject qualified ones. They are undoubtedly a consequence of the process of development of these tests, which was carried out on samples of recruits containing relatively few Blacks

or females. The technology of test construction is, regrettably, still so empirical that good psychometric properties cannot be guaranteed without pretesting the items in samples from the relevant subpopulations. It is especially difficult to make a fixed-length test that is equally reliable in groups with widely separated mean levels. Bock and Mislevy (1981) estimate the reliability of the Auto and Shop Information tests, which show the largest sex difference in mean scores, to be .83 in males and only .69 in females. Similarly, Mathematics Knowledge, which shows some of the largest sociocultural differences, has reliability .85 in White males aged 19 through 20, whereas the reliabilities in the corresponding Black and Hispanic groups are .71 and .76 respectively.

The results of Valentine's study, reviewed here, seem to indicate that if the problems of differential reliability are solved by improved tests, there will be no differential validity of the ASVAB tests, either for sociocultural group or for sex. His results seem even more favorable in this connection than those of Maier and Fuchs based on earlier tests. There is practically no evidence in either of these studies of bias in favor of Whites or of males.

### 3.5 Description of ASVAB 8

With the exception of Numerical Operations and Coding Speed, the ASVAB tests are "power" tests--that is, tests with time limits sufficiently generous to allow most examinees to attempt all items. Differences in test scores result primarily from numbers of right and wrong responses due to differences in the difficulty of the items and not from differences in the number of items attempted. Generally speaking, the items become progressively harder from the beginning to the end within each power test. Figure 3.5-1a, b, c, and d conveys something of the range of item content and difficulty within each of

these tests. Easy, medium, and hard items are shown in their respective locations on the measurement scale for each test. The units on these scales have been set so that the mean is 500 and the standard deviation 100 in the total sample correctly weighted to reflect the population composition. Each item is represented at the point on the scale where the probability is 50 percent that persons with that score will correctly answer the item. (These points are called the item 50-percent thresholds.) Because the content of the ASVAB Form 8 must be kept secure so that it can be used operationally and for standardizing other forms, we show only simulated items in Figure 3.5-1: they are similar to items at these thresholds, but words and other details have been altered to disguise the item and the location of the correct response. They nevertheless provide an indication of the progression of the item difficulty along the scale.

The speeded tests, Numerical Operations and Coding Speed, are well represented by the practice items from ASVAB Form 8A shown in Figure 3.5-2. All items in these sets are about equal in difficulty and cannot be scaled in the manner of power-test items.

In the construction of vocational aptitude batteries, the goal has been to find a relatively small number of tests, administered in not more than three or four hours, that have useful predictive validity for a range of occupations and jobs. Because test constructors are under great pressure to keep the battery short and simple, they are interested in finding tests that relate to very general dimensions of skill required in varying amounts in many or most occupations. At the same time, they may need to assess very particular areas of information that are required for specific jobs.

The approach to finding general dimensions of vocationally relevant skills that should be represented in such tests has been greatly influenced by

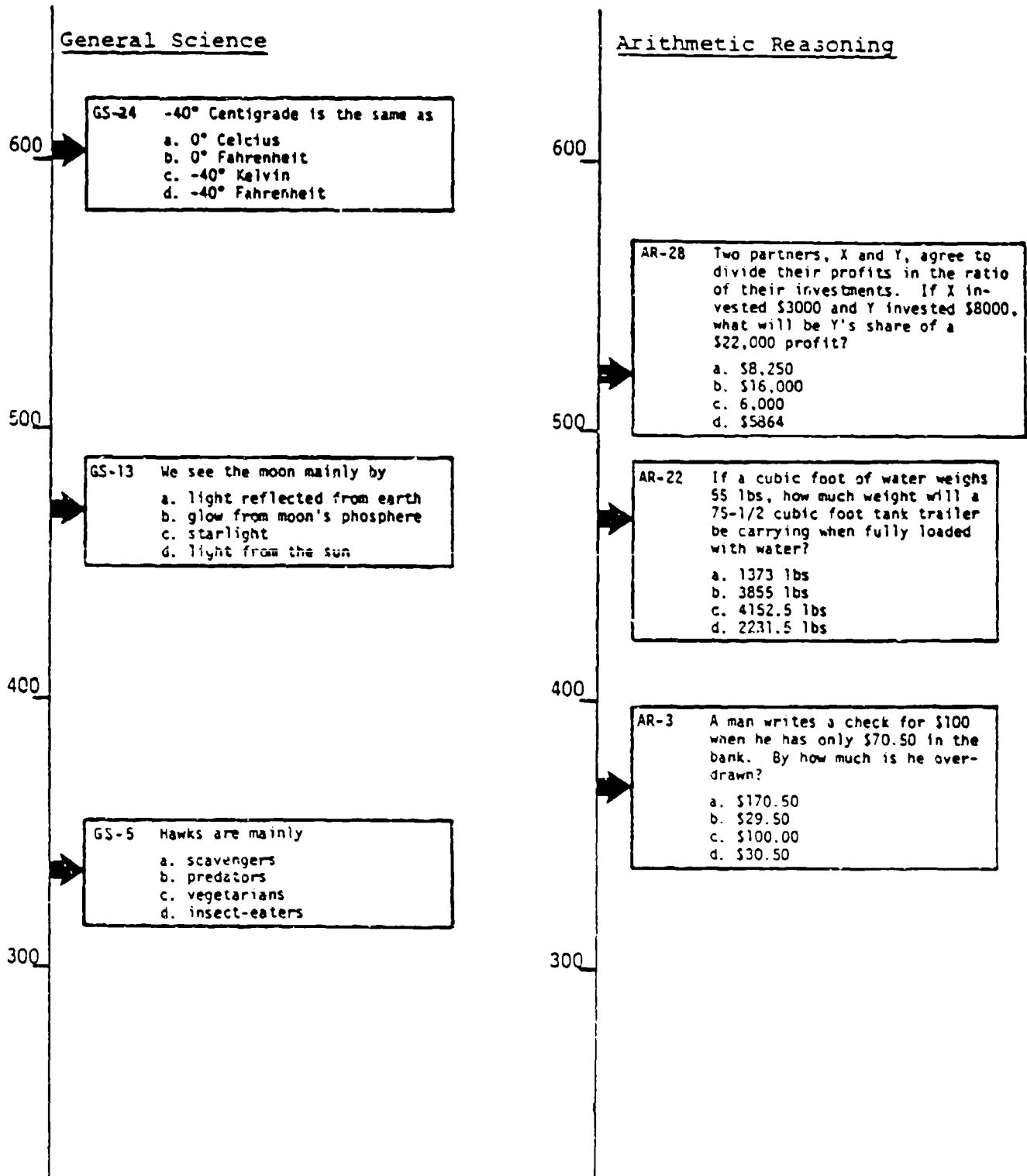


Figure 3.5-1a General Science and Arithmetic Reasoning:  
Selected item difficulties (higher scale numbers  
indicate greater difficulty)

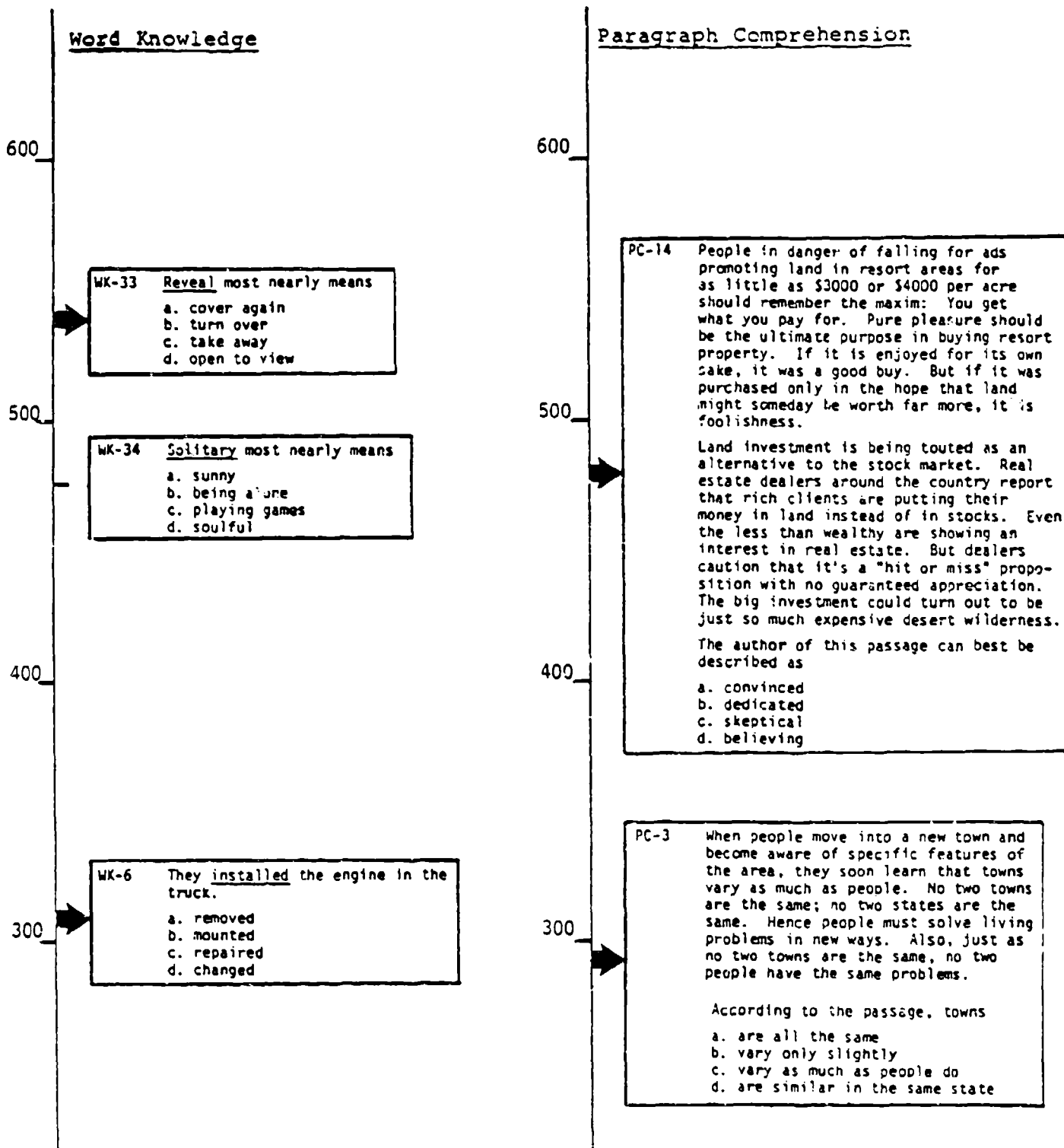


Figure 3.5-1b Word Knowledge and Paragraph Comprehension:  
Selected item difficulties (higher scale  
numbers indicate greater difficulty)

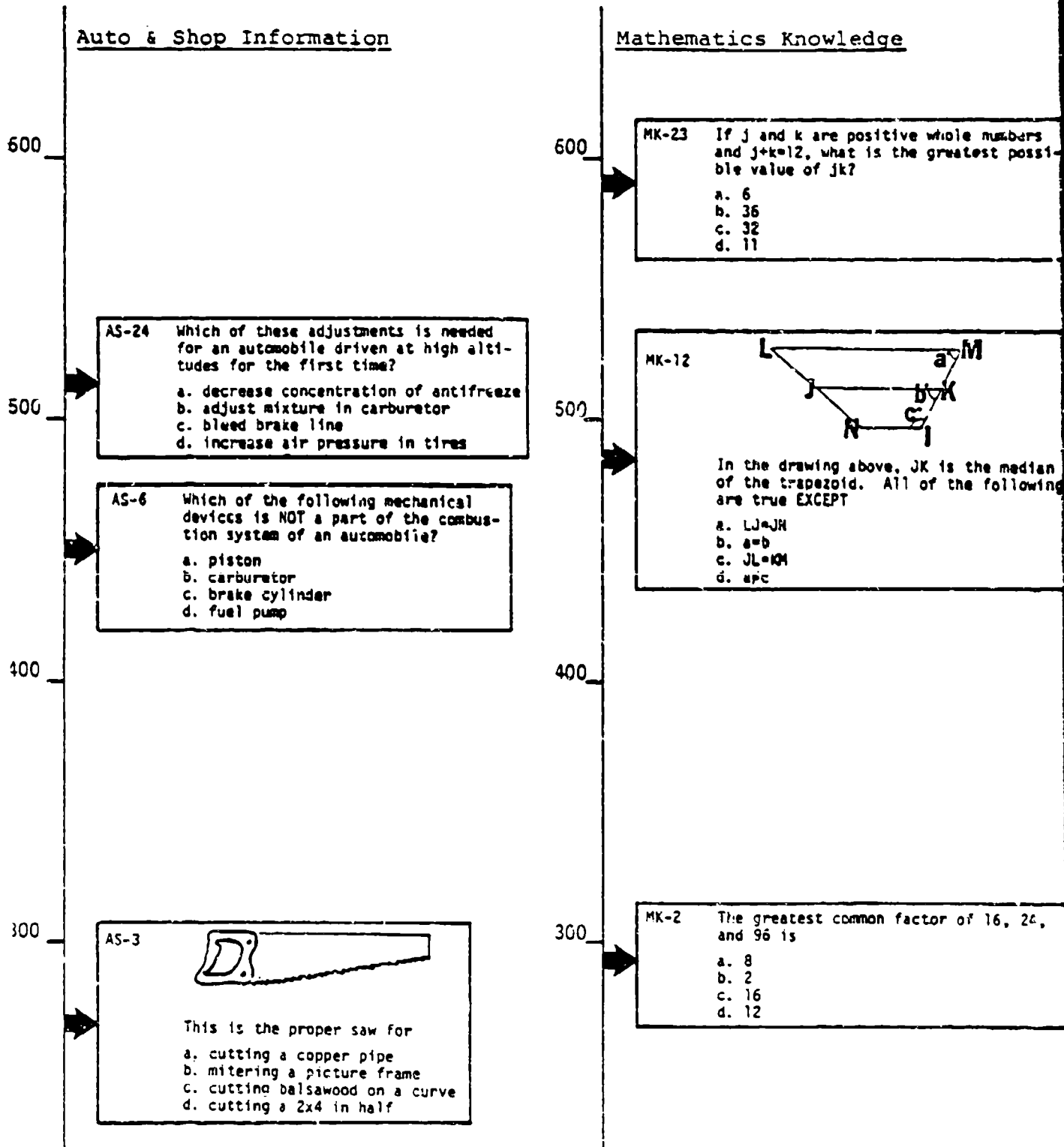


Figure 3.5-1c Auto & Shop Information and Mathematics Knowledge: Selected item difficulties (higher scale numbers indicate greater difficulty)

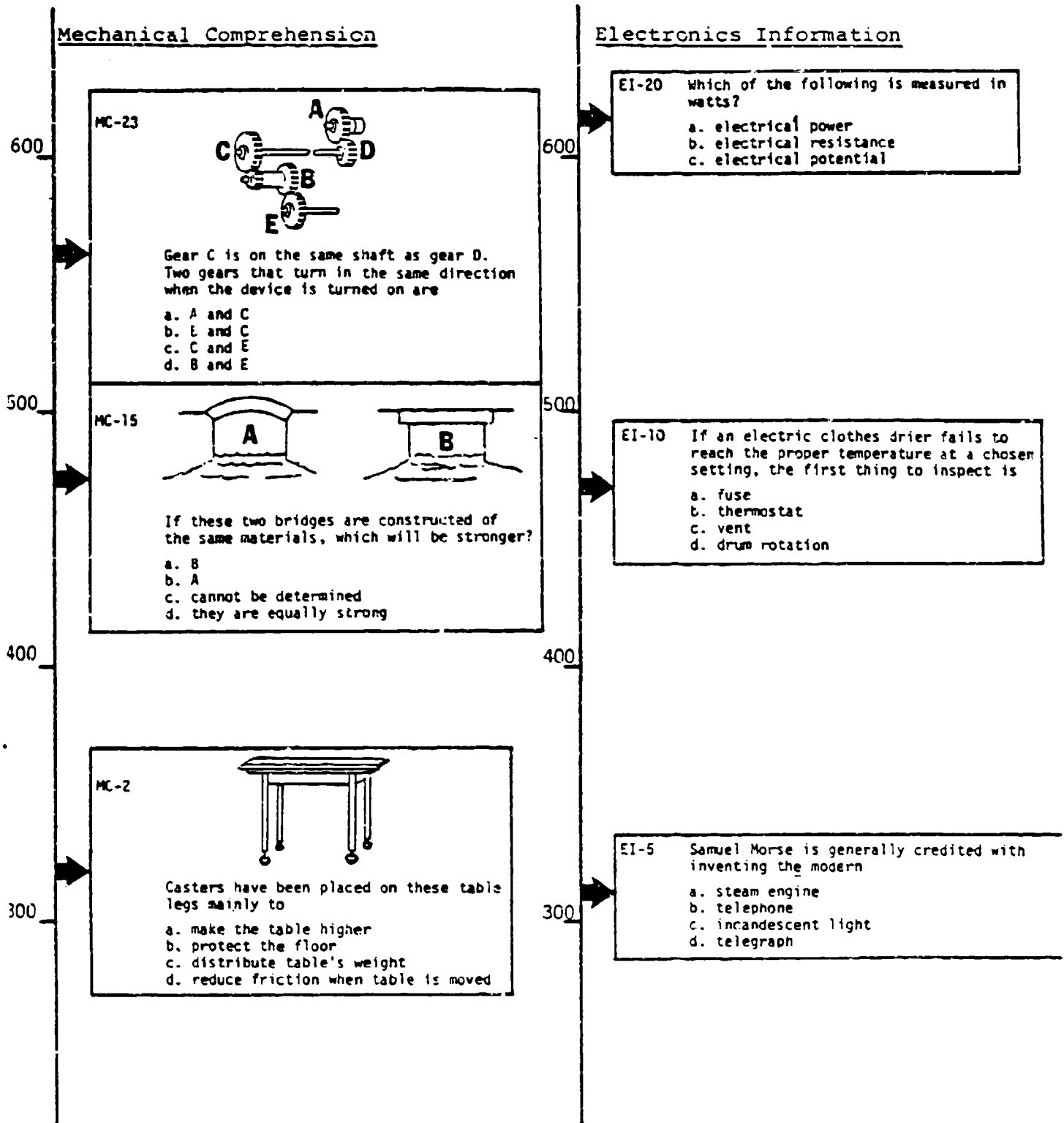


Figure 3.5-1d Mechanical Comprehension and Electronics Information: Selected item difficulties (higher scale numbers indicate greater difficulty)

### Numerical Operations

Now look at the sample problem below.

3 X 3

A 6  
B 0  
C 9  
D 1

The answer is 9, so the C answer is correct.

### Coding Speed

Look at the practice Key and the five sample questions below.

#### KEY

green . . . . .	2715	man . . . . .	3451	salt . . . . .	4586
hat . . . . .	1413	room . . . . .	2864	tree . . . . .	5972

#### SAMPLE QUESTIONS

#### ANSWERS

	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>
room	1413	2715	2864	3451	4586
green	2715	2864	3451	4586	5972
tree	1413	2715	3451	4586	5972
hat	1413	2715	3451	4586	5972
salt	1413	2864	3451	4586	5972

Figure 3.5-2      Sample problems: Numerical Operations and Coding Speed



the thinking of L. L. Thurstone (1887-1955), who with Truman Kelley (1884-1961) introduced the multiple factor analysis for this purpose. With the help of factor analysis, dimensions or clusters of related tests can be identified in greater or lesser detail. The dimensions or clusters correspond to distinct abilities in which the subjects vary, and which can be reliably measured by two or more tests. Many such systems for classifying factorially identified abilities exist (Thurstone and Thurstone, 1941; Guilford and Hoepfner, 1971; Cattell, 1967). Although these systems differ in particulars, at higher levels they identify much the same skills, although sometimes with different terminology. Most factor analysts would agree at least on the importance of the three major sources of variation in mental test scores discussed in Bock (1973)--namely, verbal ability, spatial ability, and fluent production. They appear in the factor analysis of many cognitive test batteries, and there is evidence from research on twins that they have a genetic component.

At an intermediate level, most factor analysts would also concur in identifying the skills related to the intellectual tools of our culture--namely, reading, arithmetic ability, spelling, and other language usage. Variation in these more basic abilities is probably due mostly to differences in educational and cultural background. At a still more detailed level, factors relating to specific content may appear depending on the particular tests used and the populations to which they are administered. Variations specific to these factors reflect a wide variety of experiential differences. From the content of the ASVAB, we can see that the tests are aimed simultaneously at several sources of variation.

The test of General Science, although certainly relevant for predicting success of further education in scientific fields, has a role in the ASVAB bot, as a test of the extent of exposure of subjects to technical

topics, and as a measure of the efficiency, capacity and accuracy of the long-term memory and recall of information. Because the military occupational specialties are becoming increasingly technical, recruits who have an interest in science and have learned basic scientific facts and concepts are at an advantage in many of the training courses and military specialties. The General Science test was constructed from what were originally separate tests of physical science and biology, and the distinction is still evident among the items. Rock and Mielew (1981) have shown that there are Profile differences between men and women in their response to these types of items: men do slightly better on the physical science items and women slightly better on the biological science items. The accuracy of the tests for classifying men and women would be improved, albeit at the cost of greater complexity of reporting, by scoring the physical science and biological science items separately.

The Arithmetic Reasoning test requires the ability to apply the arithmetic operations of addition, subtraction, multiplication, and division of integers and fractions. Mental arithmetic is not required; the subjects are given scratch paper and told that they may use it in the calculations. But the test is not primarily a measure of ability merely to perform these formal operations. Rather, the items require the subject to recognize the correct operation to apply in the various situations set by the problem.

The items shown in Figure 3.5-1a are typical: most are rate, pro rata, and ratio problems. Eleven of the items deal with money, and another eleven with physical dimensions; five have some connection with homemaking (e.g., extending recipes); only one contains a military reference; two are unclassifiable. The four multiple choice alternatives of each item are always numbers with the units of measurement indicated. Wrong alternatives represent

obvious types of conceptual or computational errors that the examinee might make.

Although men tend to score higher on the Arithmetic Reasoning test than women, Bock and Mislevy (1981) found no evidence of differential sex bias among these items. The items that seem directed more toward women are not actually any easier for them relative to the items with physical and mechanical content. The only apparent trend is that as the items become more difficult the proportions of correct responses of the male subjects increases relative to that of females. Either the female subjects make errors on the more difficult items, or they begin but do not complete, or do not attempt to answer, the more difficult items that appear later in the test.

Word Knowledge is a straightforward vocabulary test. The items, like those in Figure 3.5-1b, consist of a sentence in which one word is set in boldface type and underscored. The subject is to choose from among four alternatives the word or phrase that "means the same, or most nearly the same" as the underlined word. Twenty-two of the thirty-five item stems take the form "so-and-so most nearly means"; the remaining items give some minimal indication of the context in which the word is used. In Bock and Mislevy's analysis of the sex and sociocultural group effects on item difficulty, the item thresholds showed a significant sex difference for at least one of the words--a military and sporting term that appeared to be relatively easier for males than for females. The sociocultural group interactions with Word Knowledge item difficulties were confined to contrasts between Whites and Hispanics, with certain words being relatively easier for each of these groups. As we discuss in Chapter 6, the word knowledge of persons who speak both English and Spanish is affected by the fact that a somewhat infrequent and unusual word in English may have a Romance language cognate that is a

relatively common word in Spanish. In fact, many of the items in Word Knowledge take the form of a stem word of Latin origin and alternatives that include the more common Anglo-Saxon equivalent. Since the bilingual Hispanic speakers are likely to know both types of words, these items are relatively easier for them. The exception is, of course, the occasional false cognates where the English meaning has become quite unrelated to the Spanish word. These words may be relatively easier for the English speakers. Most of the English-Hispanic effects on the difficulty threshold for the Word Knowledge items can be explained in these terms.

Paragraph Comprehension is a conventional reading test in which the examinee is presented a brief passage and asked one or more questions about it. In the field of reading instruction, it is now considered important to distinguish between literal reading ability and reading comprehension. An item measuring literal reading should ask only about the facts and relationships actually mentioned in the paragraph. In reading comprehension items, the examinee may be required to generalize the theme of the paragraph or infer relationships implied by it. This latter kind of reading ability obviously places a greater premium on the experiential background of the subject. A good test of whether or not an item is a measure of literal reading is the extent to which it can be answered without reading the paragraph. If it can be so answered, it is drawing on general information and reasoning ability in addition to reading ability. In this respect, the items in the ASVAB lean rather heavily on general knowledge. A well-informed person has a good chance of answering many of the items correctly without reading the paragraph. This means that the better educated among the Profile subjects, having both the benefit of reading the passages and already knowing many of the facts in the passages, should have found this test very easy. That this is the case is evident in the distributions of scores shown in Section 4.2.

The Numerical Operations test measures the speed with which the subject can perform the simplest overlearned arithmetic operations on whole numbers, in most cases with single digits, and mark the correct alternative on the answer sheet. The time limit is so stringent that very few subjects even in the most educated groups are able to respond to all fifty of these items in the time allowed. A practice item from this test is shown in Figure 3.5-2.

Coding Speed is similar in many ways to Numerical Operations but requires no arithmetic knowledge and draws more heavily on short-term memory. Typical examples of the Coding Speed items are those shown in Figure 3.5-2. Subjects are given a key that associates a common word with a four-digit number; they are then required to mark the number corresponding to the word in a series of twenty-eight random presentations of the words. The key then changes and another twenty-eight presentations of the words follow, then the key changes a third time and twenty-eight more presentations follow. The short time limit for the test leaves little probability of anyone's completing all eighty-four items.

Tests similar to Coding Speed, sometimes called "Clerical Speed," appear in many vocational aptitude batteries. They are believed to predict the speed and exactness with which clerks or secretaries can process information from one list to another, but the ability required by the test seems more general. Because the scores on this type of test correlate fairly strongly with general verbal ability, the test of this type called "Digit-symbol" appears in the Wechsler Adult Intelligence Scale. Why this test should correlate with general intelligence is not well understood, but studies by Hunt (1976) and others seem to indicate that persons who perform well in general information and verbal tests are able to process the semantic information in symbols more rapidly than those who do less well. Such persons

undoubtedly spend more of their time in working with written symbols and language and overlearn more thoroughly the basic processes of identifying symbols and retaining words and numbers in short term memory. They can therefore perform the task of comparing symbols and transferring information between the test booklet and the answer form more quickly than persons with less facility in processing this type of material. Another factor that must enter into performance on this test is the motivation of the examinee. Both the Numerical Operations and the Coding Speed tests are fatiguing, not to say boring, and put a premium on the concentration, persistence, and enthusiasm of the subject. Those persons who can marshall these resources in the rather artificial task set by the Coding Speed test have qualities that will serve them well in many practical work situations. It is perhaps then not surprising that tests of this type have reasonably good predictive validity for a wide range of clerical occupations.

The remaining tests in the battery appraise knowledge more specifically required in particular vocations.

Auto and Shop Information, the items of which are similar to those in Figure 3.5-1c, deals with the everyday facts that one encounters in automobile repair and in metal and wood shops. Of the twenty-five items on the test, fourteen require knowledge of automobiles, eight of metal shop practice and tools, and four of woodshop tools and practice. This knowledge could readily be acquired in high school shop classes, on the job, or by reading magazines such as Popular Science or Popular Mechanics. It is not at all the kind of knowledge that would be covered in liberal arts undergraduate education.

Bock and Mislevy (1981) find considerable sex differences in the difficulties of some of these items. Male subjects tend to find all of these items much easier than do female subjects, especially items about the working

mechanisms of automobiles. Female subjects do slightly better on items that can be answered on the basis of general principles rather than specific facts of auto and shop work.

The Mathematics Knowledge test is the most formally demanding of the ten tests. Unlike the Arithmetic Reasoning test, it does not require the solution of "word problems." All twenty-five of the items are similar to those shown in Figure 3.5-1c in requiring knowledge of mathematical terms and concepts and the ability to perform formal algebraic operations. The content of these items would be known only to persons who had some exposure to high school courses in algebra and geometry or who had studied text books on the subjects.

The Mechanical Comprehension test appraises knowledge of the properties of materials and structures and how mechanical devices operate. This knowledge could come from direct experience with machines, simple and complex, from formal study of mechanics, and from reading on technical subjects. Many of the items require the examinee to interpret pictures of mechanisms which, although relatively crude, are in the style of mechanical drawings. Exposure to mechanical drawing and drafting would assist a person in responding to these items. A number of these drawings depict the movement of parts which the examinee could visualize in order to solve the problems. Equally important, however, is a familiarity with the conventions of schematic drawings of the type commonly seen in elementary physics texts and handbooks of shop practice.

The final test, Electronics Information, consists of twenty items asking simple questions about electricity and about electronics. Most of the

questions require only practical knowledge of the subject but a few of the more difficult, such as that shown in Figure 3.5-1d, draw on knowledge of principles that would have to be learned in formal study.

### 3.6 Psychometric Properties of the ASVAB Tests

Reliability. Multiple-choice tests such as those of the ASVAB are by nature fallible measuring instruments and their reliabilities, defined as the ratio of the non-error variance to the total score variance in the population of subjects, are rarely if ever as high as those of physical measuring instruments. In general, the reliability of an objective test is a function of the number of items--longer tests are more reliable than shorter tests. But longer tests require more time to administer, and in a multitest battery, the test developer must compromise between the number of tests that can be administered in an allotted time and the reliability that can be attained for each separate test. In the case of the ASVAB, the separate tests were kept rather short and the reliability compromised somewhat relative to other vocational aptitude batteries. Table 3.6-1, taken from Bock and Mislevy (1981), compares reliabilities for the ASVAB tests with those of some of the DAT tests. The reliabilities of the ASVAB tests are reasonably high, but not as high as those of the DAT tests. In operational use, however, various composites such as the AFQT composite  $[WK+PC+AR+(1/2)NO]$ , are employed in selection and classification, and these have reliabilities in excess of .90.



TABLE 3.6-1  
RELIABILITIES OF THE ASVAB AND THE DAT

Test	Number of Items	Reliability <sup>1</sup>
ASVAB:		
General Science	25	.86
Arithmetic Reasoning	30	.87
Word Knowledge	35	.86
Paragraph Comprehension	15	.68
Numerical Operations	50	.71
Coding Speed	84	.82
Auto and Shop Knowledge	25	.83
Mathematics Knowledge	25	.84
Mechanical Comprehension	25	.83
Electronics Information	20	.80
DAT:		
Verbal Ability	50	.95
Numerical Ability	40	.93
Mechanical Ability	70	.91
Spatial Ability	60	.95
Spelling	100	.96
Language Usage	60	.91

<sup>1</sup>Complement of the relative average error variance. See Bock and Mislevy (1981).

Test Intercorrelations. Although the statistical analysis in the present study is directed toward differences among groups and subgroups defined in the classification of subjects by background factors, some information about variation within the cells of the design results as a byproduct. Table 3.6-2 shows the pooled estimates of within-cell standard deviations and correlations of the ASVAB subtests from a weighted multivariate analysis of variance. (The MULTIVARIANCE program was modified especially for this purpose.) (In this Table, the tests are ordered according to their content rather than their order of administration. We will use this order in the remainder of the text.) On the assumption that the variance and covariance of the scores are homogeneous within cells, the estimates in Table 3.6-2 are

obtained by pooling weighted sums of squares and products of deviations from the cell means. The weights used are those of the national probability sample. Thus each subject contributes to the assessment of variation and covariation in proportion to his or her weight in the probability sample.

TABLE 3.6-2  
POOLED WITHIN-CELL CORRELATIONS (WEIGHTED) OF ASVAB SUBTESTS

Correl- ations	1 WK	2 PC	3 AR	4 NO	5 CS	6 GS	7 MK	8 MC	9 EI	10 AS
1 WK	1.000									
2 PC	.623	1.000								
3 AR	.557	.539	1.000							
4 NO	.374	.395	.487	1.000						
5 CS	.321	.345	.386	.589	1.000					
6 GS	.681	.532	.557	.320	.277	1.000				
7 MK	.538	.523	.745	.496	.399	.555	1.000			
8 MC	.475	.450	.547	.273	.263	.539	.493	1.000		
9 EI	.573	.468	.477	.267	.233	.615	.443	.553	1.000	
10 AS	.394	.327	.323	.177	.172	.435	.248	.525	.527	1.000
Standard Deviation	71.8	77.7	77.7	81.4	80.6	74.6	78.6	74.8	72.4	67.3

The figures in Table 3.6-2 exclude all variation due to differences between the cells of the demographic classification. The correlations therefore describe relationships among individual differences within these subclasses. Covariation due to broader social influences acting on classes as a whole has been eliminated. The latter effects are discussed as interactions in Chapter 5, 6 and 7.

Because all of the tests have been scaled with standard deviation 100 in the population, it is apparent from the range of the standard deviations between 70 and 80 in Table 3.6-2 that considerable variation remains after the demographic effects have been removed. In terms of variances, about 50 to 65 percent of the total variance in the national probability sample is due to individual differences within the cells of the background classification and

50 to 35 percent is attributable to variation between cells. Background factors contribute most to Auto and Shop Information, which is obviously highly experiential knowledge, and least to Numerical Operations, a largely overlearned skill.

To conceptualize sources of individual difference variation, it is helpful to perform an exploratory factor analysis on the correlations in Table 3.6-2. We have done this by the maximum likelihood method using the EFAP II program of Joreskog and Sorbom (1978). Assuming simple random sampling, this method provides a statistical test of the number of factors if the degrees of freedom for estimating the correlations are specified. In the present application to a weighted probability sample, it is difficult to know the actual effective sample size on which these degrees of freedom depend. To be conservative, we will assume a design effect of 2 and take the effective sample size to be half the number of subjects. That is, we will use 5,453 as the effective sample size (see Section 4.2). Subtracting 766, which is the number of nonvacant cells in the design, gives 4,687 degrees of freedom for calculating the chi-square statistics for goodness of fit after 1, 2, 3, 4 and 5 factors have been inserted in the factor analysis model. Because the program assumes a correction to the general mean, the effective number of degrees of freedom is 4,688. The results of the tests of fit are shown in Table 3.6-3.

TABLE 3.6-3  
LIKELIHOOD RATIO CHI-SQUARE FOR FIT OF THE FACTOR ANALYTIC MODEL  
ASSUMING 1, 2, 3, 4 AND 5 FACTORS (N = 4,688)

Number of Factors	Chi-square	df
1	4,238	35
2	1,866	26
3	863	18
4	90	11
5	17	5

Strictly speaking, Table 3.6-3 would indicate that five factors are required. But the fifth factor is a singleton involving only General Science, and its introduction produces a negative correlation in the promax rotated factor pattern. The four factor solution therefore seems more satisfactory, and it is also suggested by the nearly ten-fold decrease in chi-square when the fourth factor is added. The factor loadings, uniqueness, and intercorrelations of the promax factors rotated from a preliminary varimax solution are shown in Table 3.6-4 (see Hendrickson & White, 1964).

TABLE 3.6-4  
PROMAX FACTOR LOADINGS AND INTERCORRELATIONS FOR THE ASVAB TEST

Test	Factor				Uniqueness
	1	2	3	4	
WK	.904	.018	-.037	.020	.188
PC	.475	.126	.151	.075	.490
AR	.033	.092	.722	.097	.271
NO	.001	.810	.067	-.022	.290
CS	.007	.687	-.008	.045	.508
GS	.491	-.060	.175	.269	.360
MK	.049	.081	.831	-.059	.225
MC	-.080	-.022	.301	.632	.403
EI	.267	-.029	.037	.553	.410
AS	-.033	.044	-.169	.842	.437
Factor Correlations					
1	1.000				
2	.469	1.000			
3	.678	.562	1.000		
4	.654	.330	.564	1.000	

It is apparent from the substantial correlations among the factors that, within the cells of the background classification, variation in overall individual levels of performance enters into all of the test scores. In addition, however, the four major dimensions in the oblique factor structure are needed to describe patterns of performance on the ten tests. Factor 1 is

clearly the ever-present verbal attainment factor with large loadings for the Word Knowledge and Paragraph Comprehension tests, some loading of the General Science test, and a still smaller loading on Electronics Information.

Factor 2 is the familiar fluent production factor typically seen in such speeded tests as Numerical Operations and Coding Speed. The fact that Numerical Operations requires knowledge of simple arithmetic does not enter as an important source of individual differences because most of the subjects have mastered this level of skill; they differ only in how fast they can perform the mental operation and mark the result on the answer sheet.

Factor 3 is clearly a quantitative attainment factor with dominant loadings on Arithmetic Reasoning and Mathematics Knowledge. Mechanical Comprehension also makes a small contribution to this factor.

Factor 4 represents special technical knowledge covered by the General Science, Mechanical Comprehension, Electronics Information, and Auto and Shop Information.

Thus, of the three basic factors found in most cognitive tests (Bock, 1973), two are well represented in the ASVAB. They are the verbal ability and fluent production factors. The third, the spatial visualizing ability factor, is not directly represented in the ASVAB tests, but may be reflected to some extent in the Mechanical Comprehension and Arithmetic Reasoning tests. Because the ASVAB is a vocational test battery and not just a cognitive test, it also includes the mathematical and technical content represented in factor 4.

It is important to understand that the within-cell factor analysis represented here systematically excludes the between-cell variation. The pattern of relationships seen in Table 3.6-4 does not therefore necessarily hold in the effects of the background factors, which are between-cell effects. For some background effects, tests on the same factor sometimes vary together

and sometimes do not. Sex differences, for example, act in the same way on the tests with large loadings on factors 2 and 4, but somewhat differently on Word Knowledge and Paragraph Comprehension in factor 1, and much differently on Arithmetic Reasoning and Mathematics Knowledge in factor 3. This is to be expected: differences in the average performance of the two sexes do not necessarily stem from the same root as the differences among persons of the same sex.

Similar differences in background effects occur between tests involved in the special knowledge factor. Auto and Shop Information differentiates social and economic groups in ways somewhat different from Mechanical Comprehension and Electronics Knowledge or General Science.

With these exceptions, the factor pattern in Table 3.6-4 provides a useful grouping of the tests for discussing many of the background factor effects and interactions. Examples appear in Chapters 5 and 6.

### 3.7 Summary

The tests administered in this study, which compose the Armed Services Vocational Aptitude Battery (ASVAB), are similar to those of other widely used vocational test batteries such as the General Aptitude Test Battery and the Differential Aptitude Tests. The ASVAB has evolved over a period of years from tests used for personnel selection and classification during World War II. Version 3A, which was used in the present study, includes tests of General Science, Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, Numerical Operations, Coding Speed, Auto and Shop Information, Mathematics Knowledge, Mechanical Comprehension, and Electronics Information.

Although the numbers of items in these tests are somewhat less than those of other well known vocational aptitude batteries, the reliabilities of the ASVAB tests are quite adequate, especially in view of the fact that they

always appear in linear composites when employed for selection or classification. A number of validity studies of version 8A and its immediate predecessors show such composites to be accurate predictors of end-of-course grades in Service training schools, as well as of job performance ratings in civilian occupations.

Studies of the differential validity of the ASVAB tests in male and female and White and Black subpopulations show no real evidence of bias, either favorable or unfavorable, to members of any of these groups. There is evidence, however, that some of the composites are less accurate in predicting the success of women in certain training schools than predicting that of men. This probably reflects the fact that the tests were developed using all male samples at a time when few women were assigned to these schools. Similarly, some of the composites are less accurate in predicting end-of-course grades for Blacks relative to their accuracy for Whites.

Factor analysis of the ASVAB test scores, after eliminating systematic effects due to the sex, education level and sociocultural group membership of the subjects in this study, reveals four clearly significant factors. The factors are correlated in the range .3 to .7, and they may be characterized as Verbal Skill, Fluent Production, Mathematical Skill, and Special Technical Knowledge. The spatial factor, which is present in other vocational test batteries, is not well represented in the ASVAB tests.

## CHAPTER 4

### METHODS OF ANALYSIS

To investigate differences in performance on the ASVAB tests among the various subgroups defined by the background classification, we use the technique of multivariate analysis of variance. This technique provides analysis of variance of the background effects and interactions for the ten ASVAB scores simultaneously and takes into account the correlations among the tests. A description of the application of multivariate analysis of variance to the types of nonexperimental data obtained in the Profile Study may be found in Chapter 6 of Bock (1975). The aim of such applications is to find the simplest linear model that will describe all significant score variation between the subclasses of the sample design. In the present study, the design is generated by the crossing of the Sex, Sociocultural Group, Mother's Education, Highest Grade Completed, Economic Status, and Region classifications. Although the design contains 960 subclasses, the distribution of cases among the subclasses is uneven and only 766 subclasses actually contain subjects. The remaining 194 subclasses are empty and not included in the fitting of the model. As in most survey data, the cell frequencies in the present study tend to be roughly proportional to the marginal frequencies; the design is therefore not grossly unbalanced.

The model fitting was carried out by the exact least squares method, using the MULTIVARIANCE program of Jeremy Finn (1978). The statistics calculated by the program enable us to choose a parsimonious model and calculate best (minimum variance) estimates of the effects included in the model. The estimated effects tell us how much various combinations of background classes can be expected to contribute to the test scores. To give a simple example,



the analysis shows that, on average, the fact that a subject is female adds 9.8 points to her Paragraph Comprehension score relative to that of a male subject who is identical on all other background factors. The fact that she has some college education adds 46.2 points to her score on this test relative to an otherwise comparable subject with high school education only. By combining such estimated effects, we obtain the statistics used in Chapters 5, 6, and 7 to display the results of the Profile survey.

To justify this structural, model-oriented approach to analysis, we must satisfy a number of statistical assumptions. These concern, on the one hand, the suitability of the area-clustered probability sample for the analysis of variance technique (which usually assumes simple random sampling) and, on the other hand, properties of the scores produced by the ASVAB tests.

#### 4.1 Sampling Considerations

To apply either univariate or multivariate analysis of variance to a complex cross-classification of survey data, the standard assumption is that within each of the ultimate subclasses (cells), the measurements or scores are normally distributed and have homogeneous variances from one subclass to another. If the measurements are multivariate, as they are in the case of the ASVAB scores in this study, the assumption is one of multivariate normality and homogeneity of variances and covariances. It is also assumed that the observations (that is, scores obtained from different subjects) are independent within the subclasses. Given the latter assumption, and assuming simple random sampling within cells, the sampling variance of the cell mean is equal to the within-cell variance of the particular measure divided by the number of observations within the cell. If we take the reciprocal of the sampling variance of the mean as a measure of the information it supplies to the analysis, then the precision of the analysis is greatest when the contribution of

each cell mean to the overall group comparisons is proportional to the number of subjects in the cell.

Unbalanced Design. Note that the analysis of variance places no restriction on the number of subjects in each cell. The requirement of equal or proportional numbers in the cells that is introduced in elementary treatments of the subjects is primarily a computational convenience. We do not need that assumption here because the MULTIVARIANCE program is capable of performing an exact analysis on unbalanced designs, even including those with no observations in some subclasses. Although it is more efficient to have equal numbers in the subclasses (in the sense that very large numbers of subjects in a few cells does not much improve overall precision and therefore wastes data), the analysis is not vitiated by the wide range of variation in the numbers of subjects in the cells of the present design. In fact, the oversampling of Hispanics, Blacks and poverty-level Whites is advantageous in improving the precision of comparisons involving those groups.

Weighted or Unweighted Observations. An important consequence of the invariance of the analysis of variance with respect to arbitrary numbers of subjects in the cells is that the weights employed in the population descriptions in Chapter 2 are not needed to correct the oversampling, provided that sampling allocations correspond to the classes or subclasses contained in the analysis of variance design. In that case, the design breaks up the sample on the same criteria used in assigning weights for the probability sample, and the weights within the cells of the analysis of variance design are uniform. This is very nearly true of the data. The only variation in weights within cells applies to certain metropolitan and nonmetropolitan census tracts, but even that variation is largely absorbed in the classification by subject's highest grade completed and mother's highest grade completed. For the

purposes of group comparisons, the weights can safely be ignored. If we have any doubt on this matter, we can perform the multivariate analysis of variance with and without the weights and compare the results. If the estimated between-group effects and interactions are essentially the same in both analyses, then any variation in weights within cells is small and can be ignored.

We have in fact performed such an analysis and find there is no appreciable difference in the between-group analysis based on weighted and unweighted data. For example, Table 4.1 compares weighted and unweighted estimates of sociocultural by economic status by education effects. Differences between the estimates are inconsequential and show no pattern that would suggest systematic bias in the weighted estimates. The assumption of homogeneous weights within the ultimate subclasses of the six-way analysis of variance design therefore appears quite justifiable.

TABLE 4.1

WEIGHTED AND UNWEIGHTED ESTIMATES OF SOCIOCULTURAL GROUP BY  
ECONOMIC STATUS BY EDUCATION EFFECTS FOR  
THE WORD KNOWLEDGE TEST

Highest Grade Completed	White		Black		Hispanic	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
Poor						
0- 8	408	404	366	364	388	391
9-11	460	450	389	391	441	443
12	517	509	419	422	486	490
13+	588	581	473	481	540	541
Non-poor						
0- 8	433	426	370	367	427	422
9-11	501	498	421	423	467	468
12	532	538	451	455	512	514
13+	581	584	500	503	549	549

Since the two results are so similar, one may ask why the unweighted analysis is performed at all. The reason is that only the unweighted analysis makes efficient use of the additional information contained in the over-sampling of poor Whites, Blacks, and Hispanics. The means for these groups are determined more accurately than would be indicated by the use of weighted proportions of subjects in these groups. Another reason for not using the weights is that they violate the assumptions of the tests of significance used in the model fitting. In the presence of the weights it would be difficult to judge whether certain interactions should be included or excluded from the model for the estimation of between-group effects.

Methodological research now underway aims to apply the weights within cells while retaining all information between cells. When these developments reach the stage of practical application, the Profile data could be reanalyzed without the assumption of homogeneous weights within cells and without loss of overall efficiency. In view of the near identity of this approach to the unweighted and inefficient weighted analysis referred to above, however, it is unlikely that the reanalyses would noticeably change the unweighted results.

Intraclass Correlation. When there is replication within the cells of the analysis of variance design, as there is in the present study, it is important to know whether or not the observations within cells are independent. If not, a correction may be required when calculating the pooled within-cell estimates of variance and covariance. Lack of independence within cells does not alter the expected mean for the cell, and hence has no bearing on the estimation of between-group effects; but it does alter the estimate of within-cell variance and so affects the statistical tests and confidence intervals. This consideration is important in the present study because the data were obtained by area cluster sampling. The factor by which the cluster

sampling reduces error variation, called the "design effect," can be estimated and is reported for various levels of aggregation of the data by Frankel and McWilliams (1981). In general, the design effect due to this type of cluster sampling is smaller for narrow subdivisions of the sample and increases as subsamples are aggregated. With a few exceptions, the largest design effects observed are about 2; that is, the sampling variance can be reduced by as much as a factor of 2 as a result of the cluster sampling.

In the present analysis, the ultimate subclasses are defined at such a detailed level that there is little apparent effect of cluster sampling on the within-cell variation. The cluster sampling effects have been largely absorbed into the high-order interaction of the background classes. The evidence for this is that the F-statistics for high-order interactions (residual) that we will see in Table 4.3-1 are very close to unity. This means that both the residual variation and the within-cell variation are estimating the same sampling error variation. This suggests that we should assume a design effect somewhere between 1 and 1.5 when interpreting the tests of significance in Table 4.3-2. And there is other evidence for the correctness of this conclusion: interactions that appeared to be only modestly above the level of nominal significance have reasonable interpretations when the corresponding interaction plot is examined. The case in point is the plot of mother's education and sociocultural group shown in Figure 6.1-3, the interaction of which has a plausible interpretation from what we know of the Hispanic family structure. If the more commonly assumed design effect of 2 were used, this interaction would not be judged significant; yet its reasonable interpretation convinces us of its reality. For these reasons we will use a figure of 1.25 times the nominal percentage points of the multivariate and univariate test statistics when judging the statistical significance of between-group effects and interactions.

Because the interpretation of the tests of hypothesis depends upon our assumptions about the design effect, we have chosen not to assign probability levels to the multivariate and univariate F-statistics in Tables 4.3-1 and 4.3-2. Instead, we give the nominal .05 and .01 points of their null distributions. These values can be multiplied by a factor to take into account the design effect when interpreting the statistics in each column of these tables. We believe that a factor of 1.25 is sufficiently conservative for this purpose. In other words, we would consider a statistic in the tables significant if it exceeds the nominal .05 or .01 point by a factor of 1.25. Note, however, that we interpret the univariate F-statistics only under the protection of the multivariate F for that effect; that is, the univariate statistics for a given effect are interpreted only if the multivariate statistic for that effect is significant. Effects that are not multivariate significant are not included in the model for purposes of estimation. (See Bock, 1975, Chapter 6, for a discussion of protected F-tests.)

#### 4.2 Data Properties

Multivariate analysis of variance makes two demands of the data that are not always easy to satisfy in nonexperimental studies. First, the data for each subject must be complete. If there are multiple response variables (such as the scores on the ten ASVAB tests), each subject must have a valid measurement on each variable. In addition, the assignment of the subjects to the classes of the analysis of variance design must be unambiguous. For the Profile study, this means that valid background information must be available for each subject and that the subjects are correctly classified.

Second, the distribution of score variation within the ultimate subclasses of the analysis of variance design is assumed to be multivariate normal and homogenous in variance and covariance from one subclass to

another. No assumption is made about the distribution of scores in the population as a whole. In fact, in application to a comparative study of group effects, the analysis of variance regards the data as arising, not from a single population, but from as many populations as there are cells in the cross-classification of the subjects. It is only the individual-difference variation within these cells that is treated as random and for which distribution assumptions must be made (see Bock, 1975, Chapter 1, for a discussion of the assumptions made in the analysis of variance of experiments, comparative studies and surveys).

In this section, we examine questions of how well the Profile study meets the data-property requirements and what effects any variation might have on the analysis of variance.

Missing Observations. The assumption of complete response measurements on each subject was met by excluding, from among the 11,914 subjects tested, thirty-six who did not complete the tests under the standard conditions for administration of the ASVAB battery. In addition, some subjects had to be excluded because their background information was missing or unclassifiable. As indicated in Chapter 2, the information on Age, Sex and Sociocultural Group was complete; no subjects were excluded for want of this information. But a total of 973 lacked valid information on subject's Highest Grade Completed, Mother's Education, Region of Residence at Age 14, or Economic Status relative to the official OMB definition of poverty.

Counted among the subjects for whom information is missing must also be the 772 among the 12,686 interviewed who declined to be tested or were otherwise lost from the initial NLS sample. If these missing or excluded 1,781 cases have the same ability distribution as the 10,905 cases in the analysis, no bias will result; the effect will be only a small reduction in

the power of the statistical tests and about a 7 percent increase in the standard errors of estimate effects.

One way to check on the characteristics of the missing subjects is to compare the estimates of the population composition based on the interview sample with those based on the subjects used in the present analysis. For example, Table 4.2 shows for these subjects the estimated population percentage for the background factors that are, as we shall see in Chapters 5 through 7, most related to levels of ASVAB test performance. This table may be compared with the similar estimates in Table 2.5-5. Differences between these tables are attributable to the subjects not tested or lacking background information.

As we see, the figures in these two tables are, with one exception, identical to about .1 percent. The exception is a slight tendency for the analysis sample to contain fewer Hispanics with not more than eight years of education and more with thirteen or more years. In other words, there is some indication that Hispanics in the analysis sample are drawn from slightly higher educational levels than those in the general population. In view of the correlation between educational level and test performance, this might suggest that those found in the Highest Grade Completed classes are of somewhat higher ability than those in these classes in the population. The differences are small, however, and are unlikely to have any important effect on the group comparisons. No such differences are found in the White and Black groups.

Score Distributions. In certain respects, the "number-right" scores in which test results are traditionally reported are not ideally suited to any form of univariate or multivariate normal statistical analysis. (Our analysis is not based on number-right scores; see below.) It cannot be assumed in



TABLE 4.2

POPULATION COMPOSITION FOR THE AGE GROUPS 15 THROUGH 22 ESTIMATED  
FROM THE WEIGHTED ANALYSIS SAMPLE: HIGHEST GRADE COMPLETED  
BY SOCIOCULTURAL GROUP BY ECONOMIC STATUS (N = 10,905)

Economic Status		White		Black		Hispanic		All	
		Poor	Non-Poor	Poor	Non-Poor	Poor	Non-Poor	Poor	Non-Poor
Highest	0- 8	14.3	4.0	13.2	5.7	20.2	8.3	14.6	4.4
Grade	9-11	42.3	44.6	59.3	46.0	53.2	51.5	49.7	45.0
Completed	12	22.7	31.3	18.2	30.4	17.3	24.4	20.4	30.9
	13+	20.8	20.1	9.3	17.9	9.3	15.7	15.3	19.7
TOTAL		100.1	100.0	100.0	100.0	100.0	99.9	100.0	100.0
Percent		7.7	72.9	5.4	8.4	1.7	4.0	14.8	85.3

general that the distributions of such scores will be normal. Number-right scores reflect the distribution of item difficulties within the test, and tests with too many easy or too many difficult items may exhibit substantially skewed distributions of test scores. In the present study, we are examining groups covering an extremely wide range of abilities and knowledge. The gap between the poor minority group sample with 8 or less years of schooling and the non-poor majority group with 1 year of college completed is enormous. It would be too much to expect the ASVAB power tests to yield equally good measurement throughout this range. Measurement with equal precision in such diverse groups could be accomplished only by adaptive testing procedures in which items were selected at difficulty levels suitable for each group.

In fact, some of the ASVAB tests are targetted to an ability range somewhat below the population mean, and thus are less well suited to measurement in the high ability group. Exceptions are the Arithmetic Reasoning and Mathematics Knowledge tests, which are somewhat more difficult. The other power tests tend to be too easy for high ability subjects, and scores in the high ability groups are skewed toward high scores. This is especially true of

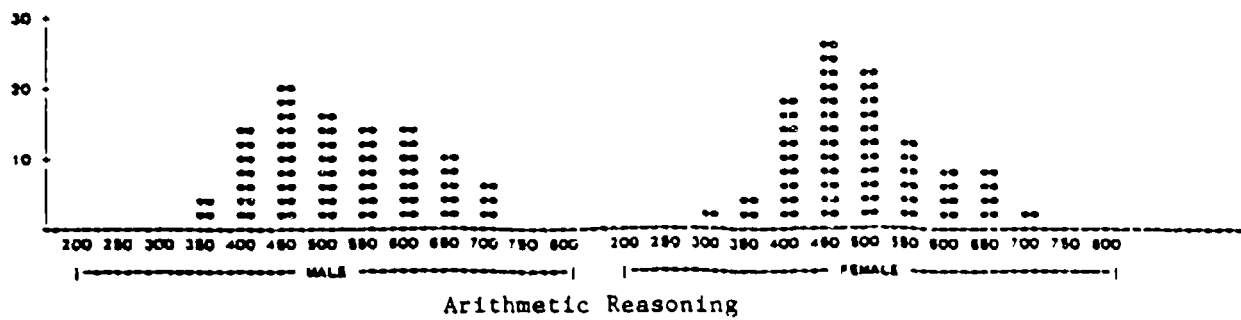
Paragraph Comprehension, which is intended only as a measure of basic reading competence and is not designed to make distinctions among high levels of reading ability. The speeded tests, Numerical Operations and Coding Speed, are capable of precise measurement throughout the entire range of ability. Even the most able subjects cannot complete all the items on either of these tests in the time allotted. For these tests, the assumption of a normal distribution of test scores within the groups is quite reasonable.

In an effort to improve the distributional properties of the power tests, we have chosen to use latent-trait estimates of ability rather than simple number-right scores. Technically, the scores we use are so-called "maximum a posteriori" (MAP) estimates, computed from item parameters estimated by Bock's fixed-effect method (Bock, 1976; Kolakowski and Bock, 1981). The three-parameter logistic response model with guessing parameter set at 0.1 was used. Details of the scoring procedure may be found in Bock and Mislevy (1981). Unlike number-right scores, which show floor and ceiling effects, latent-trait estimates tend to be normally distributed. The improvement is not so great as to remove entirely the skew toward high values in the high ability groups, but the scores on the trait scale are better in this respect than the number-right scores.

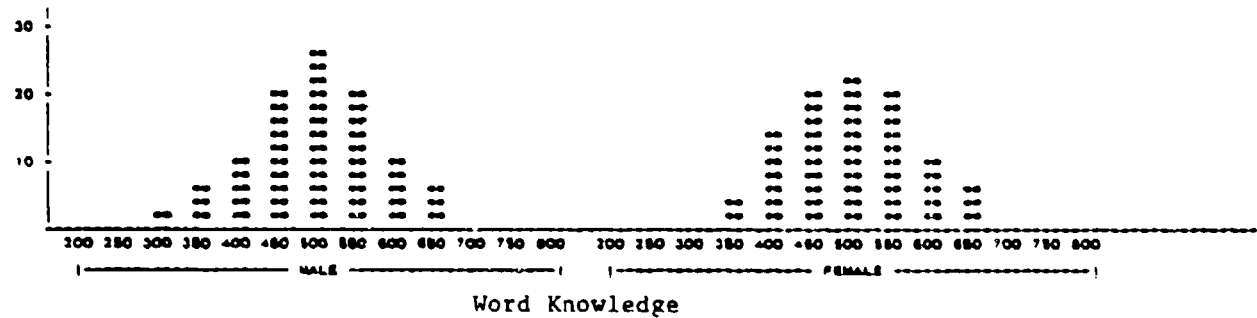
Some indication of the shape of the score distributions is conveyed by the histograms in Figures 4.2-1 through 4.2-9. These histograms describe the distributions for Arithmetic Reasoning, Word Knowledge, Numerical Operations, and Auto and Shop Information in certain of the sex by sociocultural group by education subclasses. The proportions shown in these figures, computed using the probability sample case weights, represent correctly the populations of persons in these subclasses in the nation as a whole. To simplify the presentation, we have restricted the education levels to those persons who

Figure 4.2-1 Whites with some high school by sex

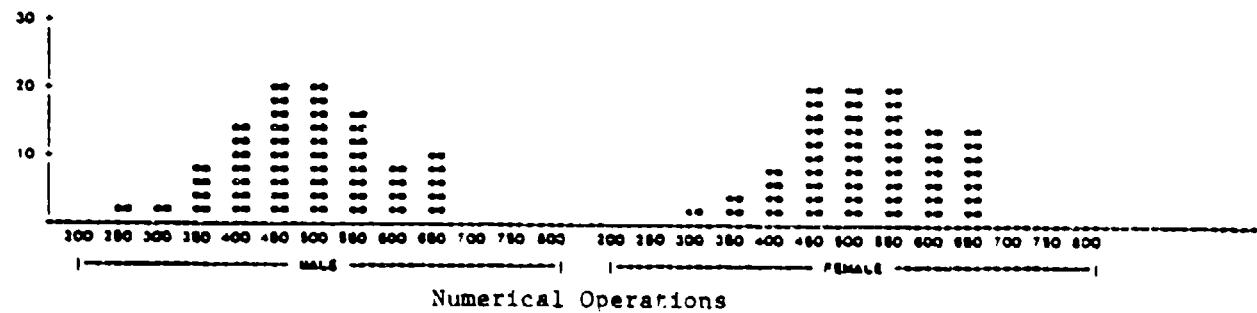
Percentage



Percentage



Percentage



Percentage

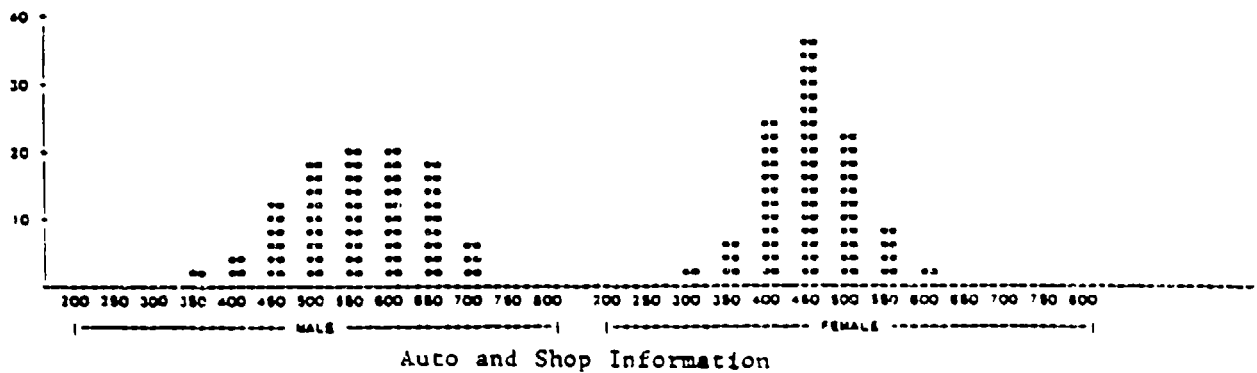
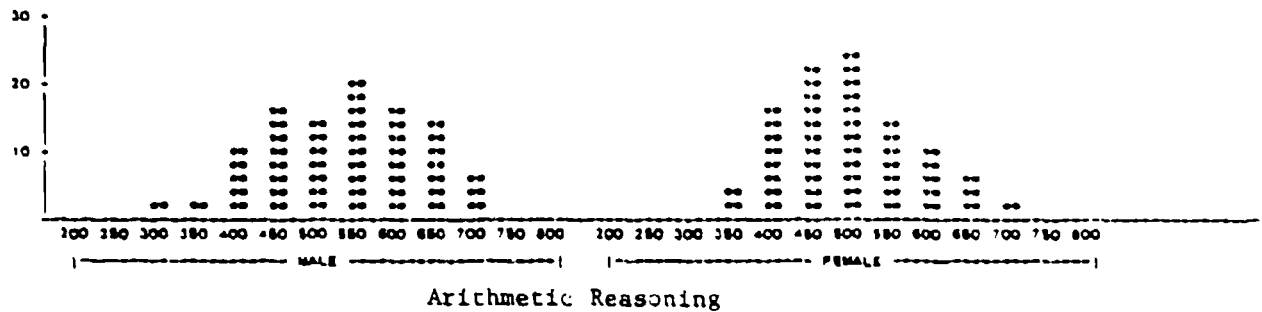
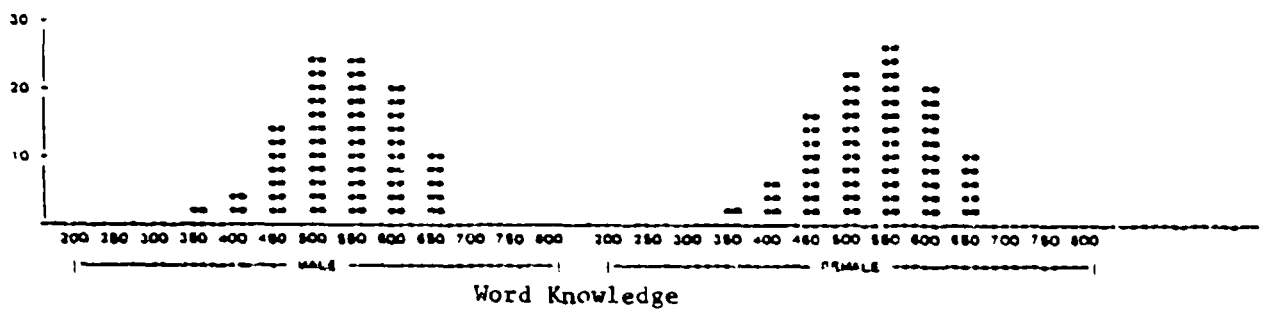


Figure 4.2-2 Whites with high school by sex.

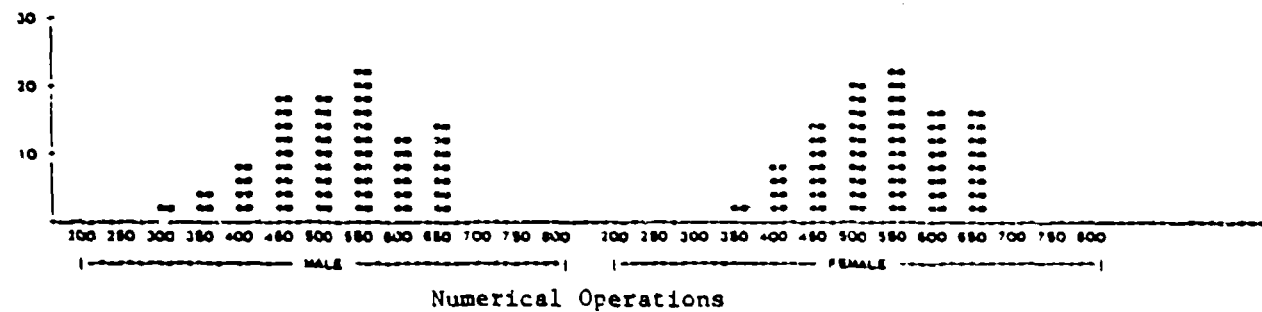
Percentage



Percentage



Percentage



Percentage

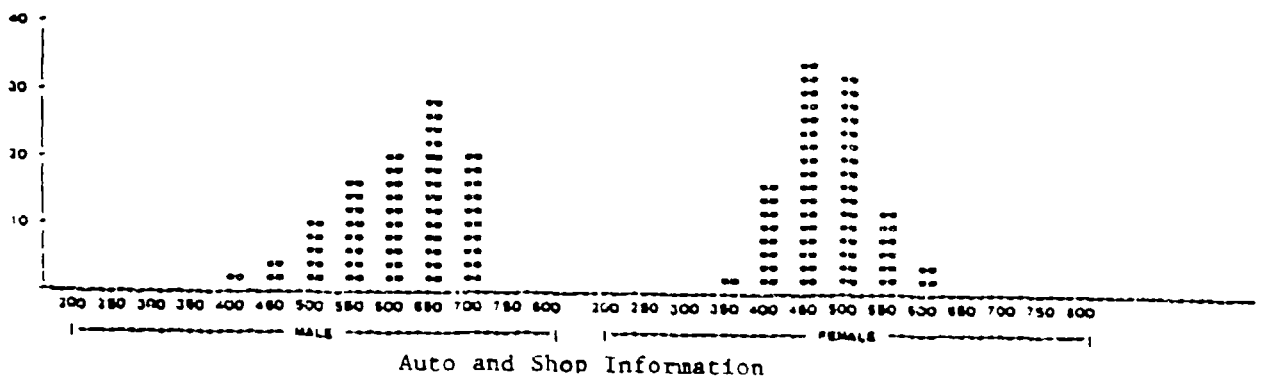


Figure 4.2-3 Whites with some college by sex (includes college graduates).

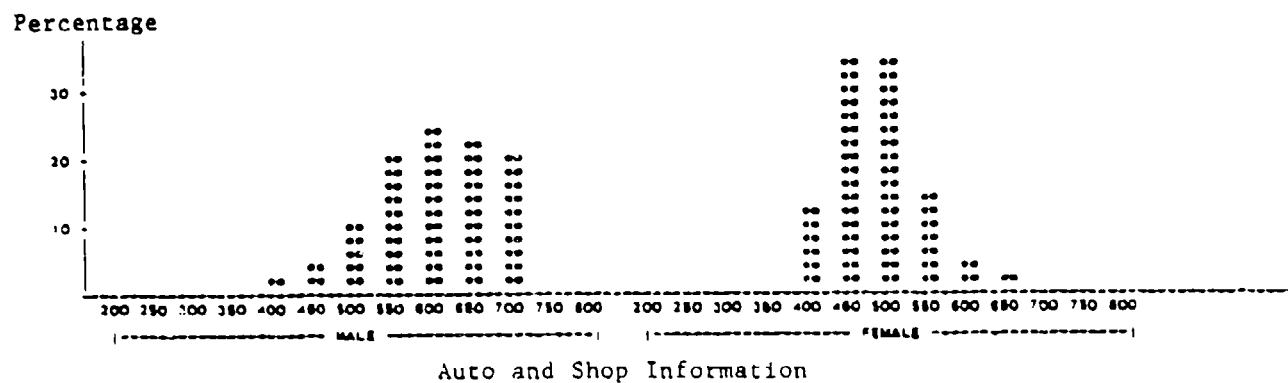
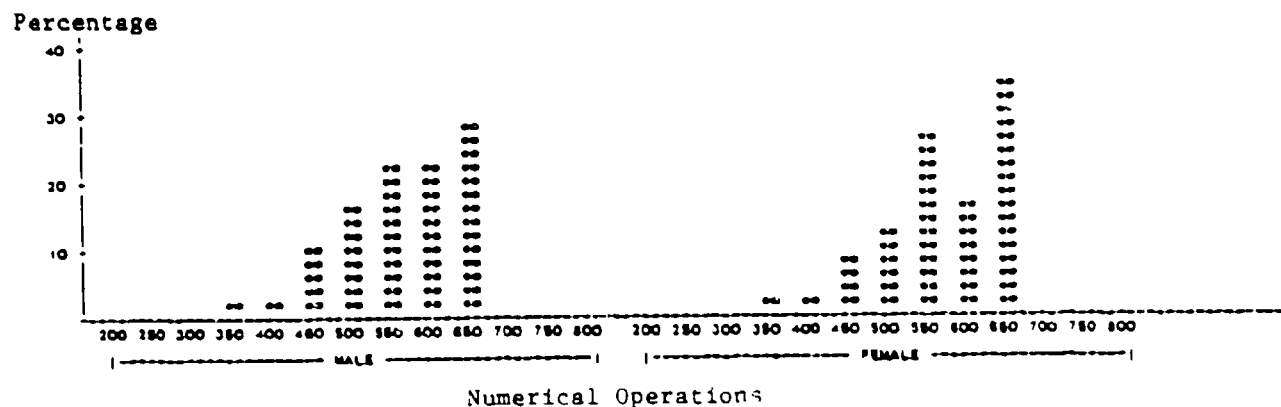
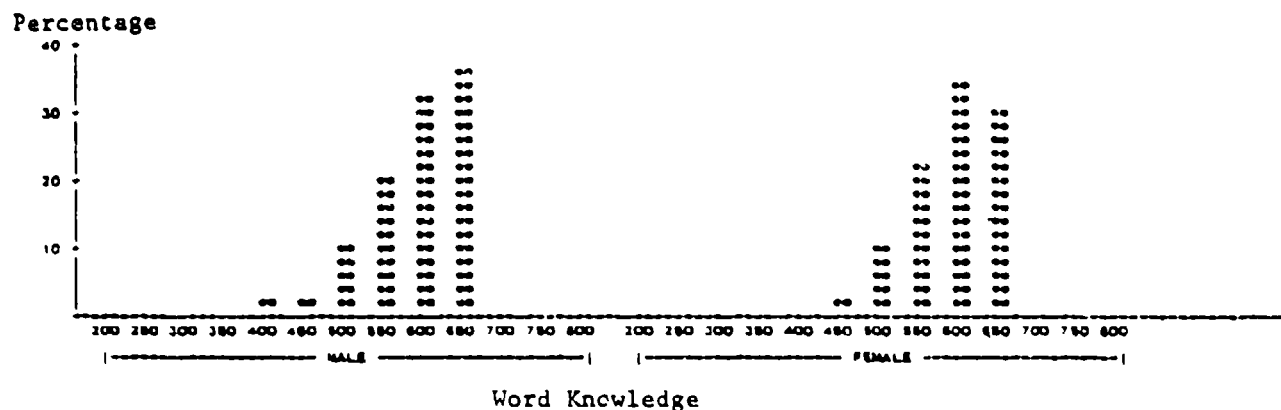
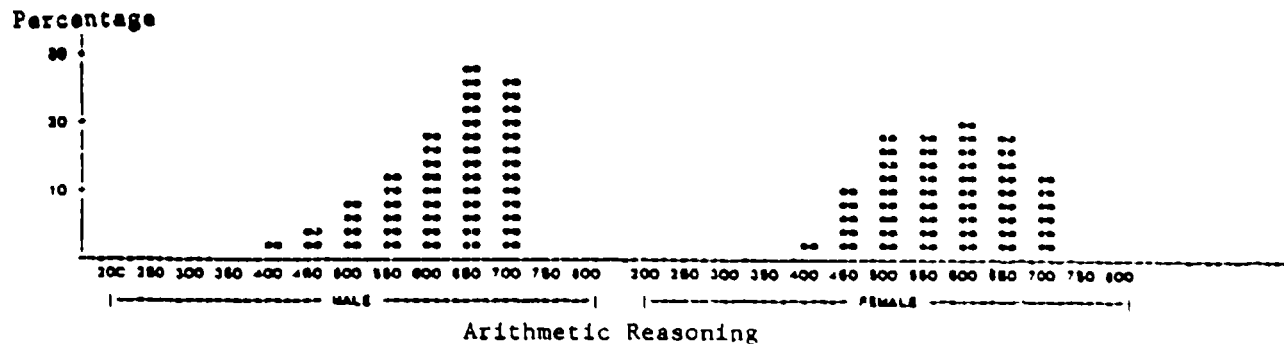


Figure 4.2-4 Blacks with some high school by sex.

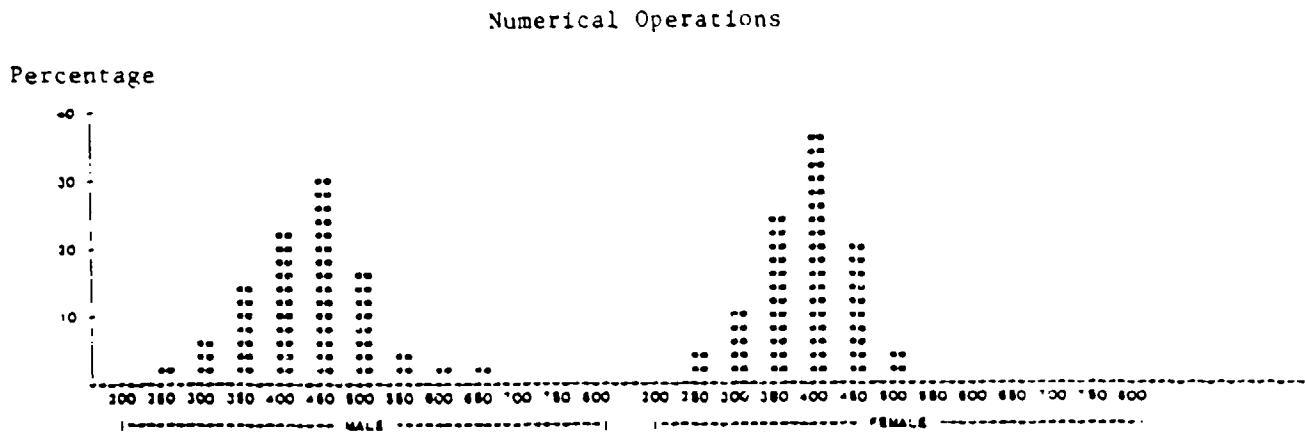
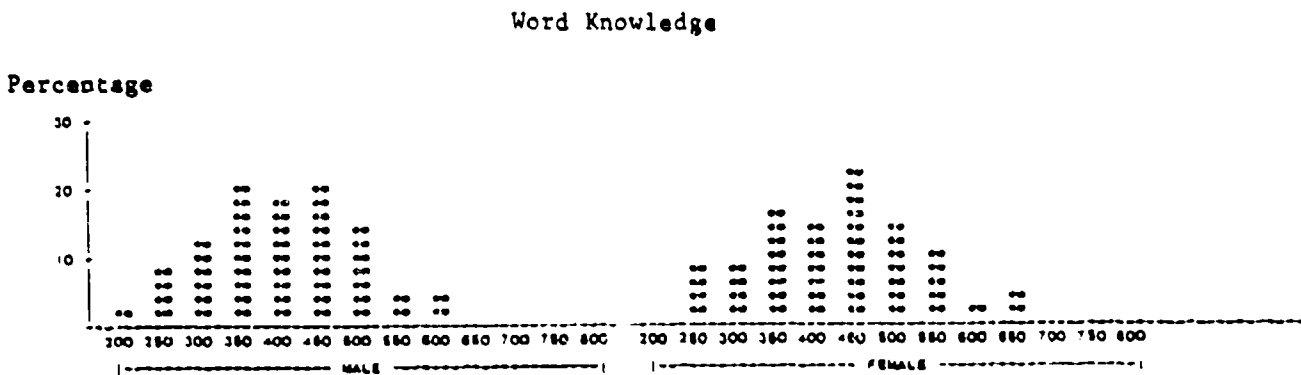
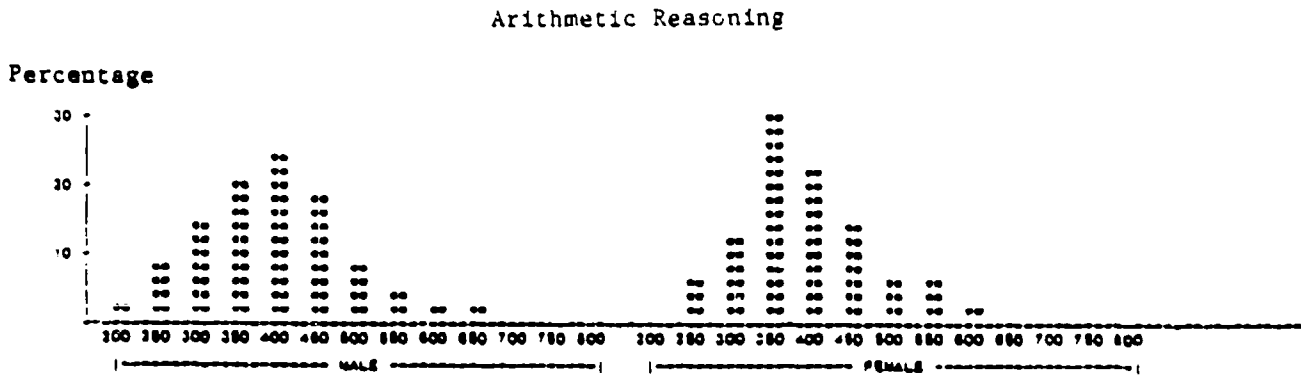
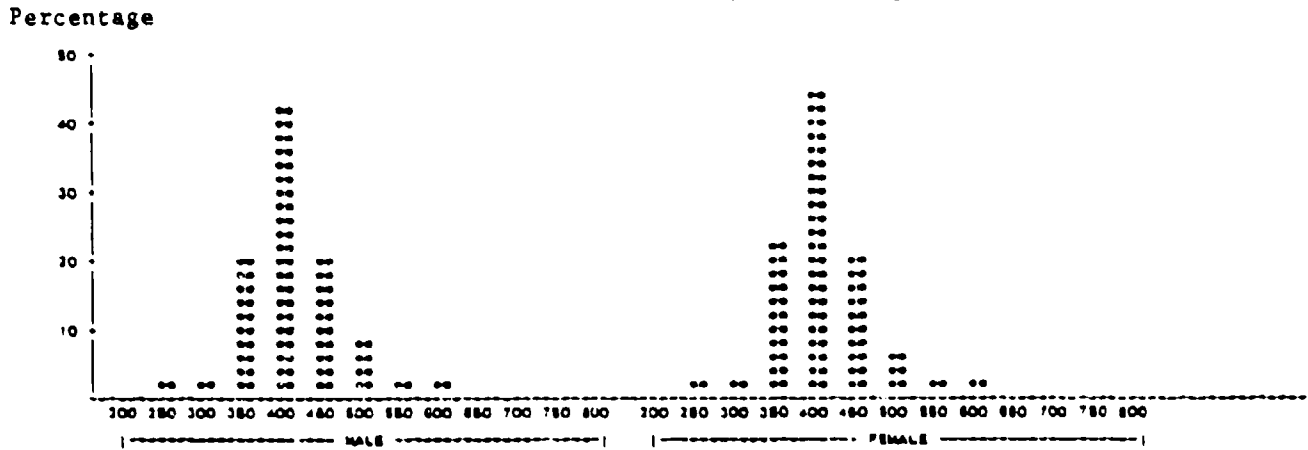
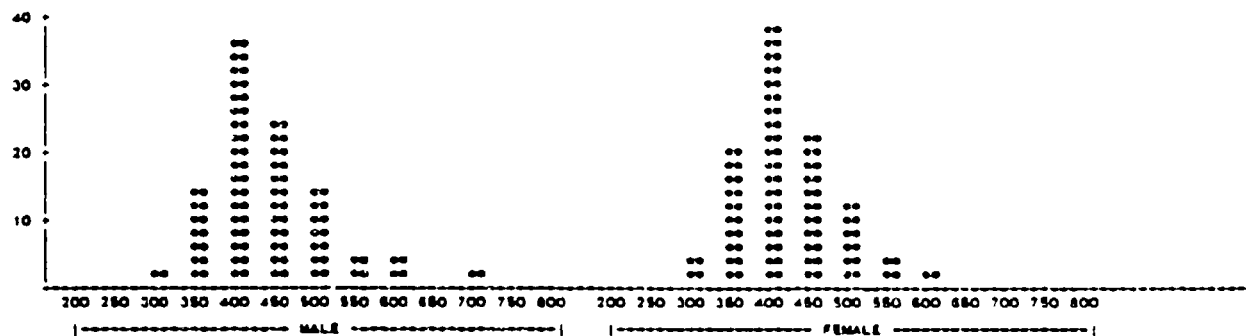


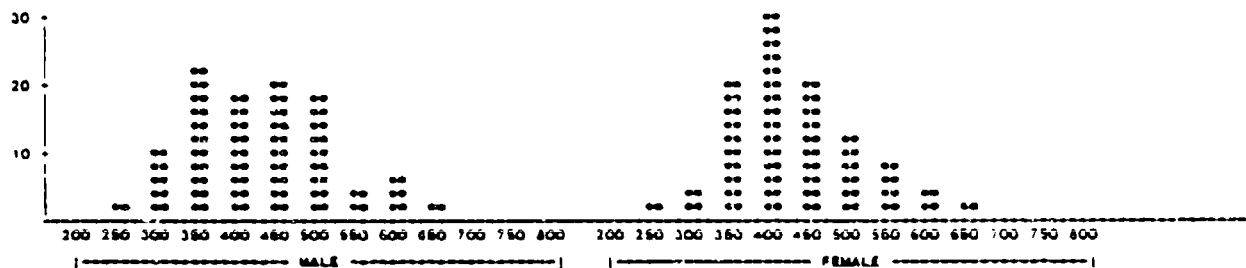
Figure 4.2-5 Blacks with high school by sex

Percentage



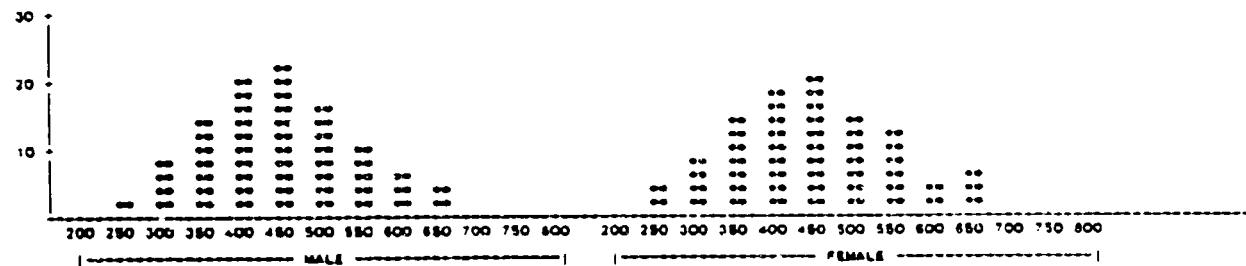
Arithmetic Reasoning

Percentage



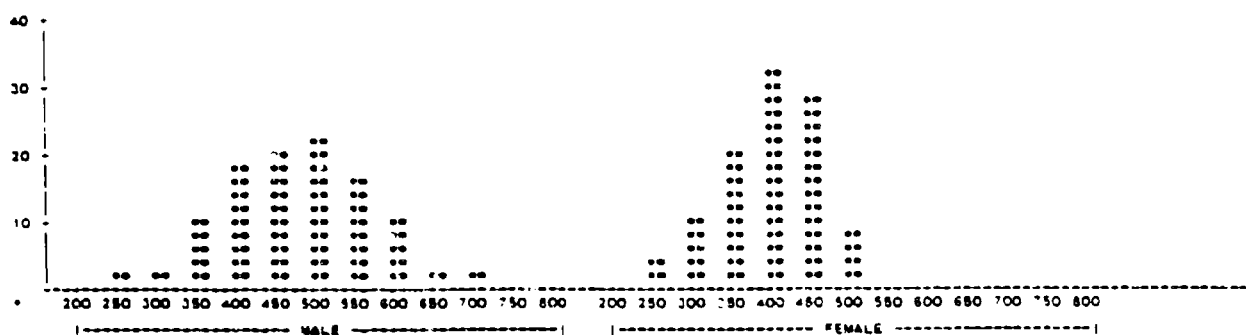
Word Knowledge

Percentage



Numerical Operations

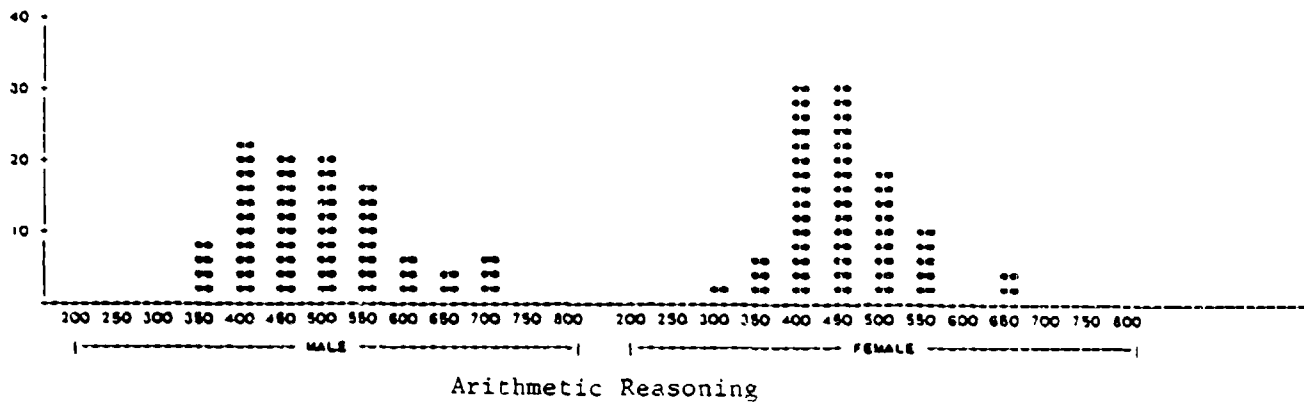
Percentage



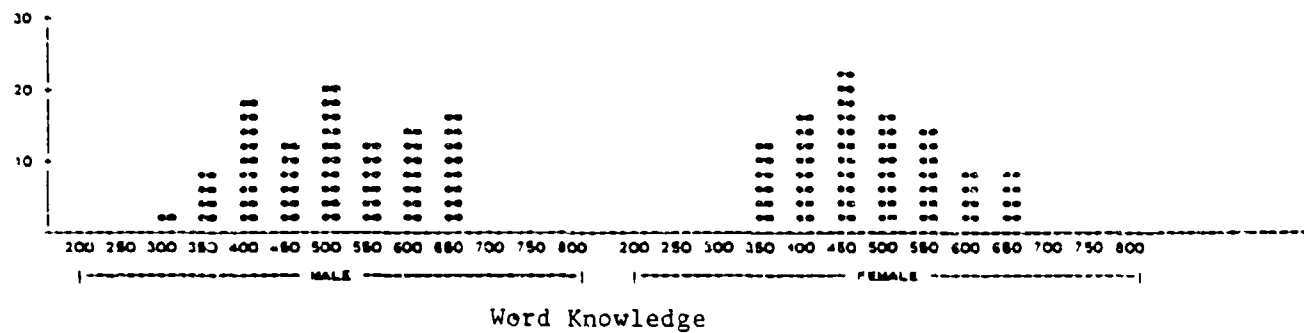
Auto and Shop Information

Figure 4.2-6 Blacks with some college by sex (includes college graduates).

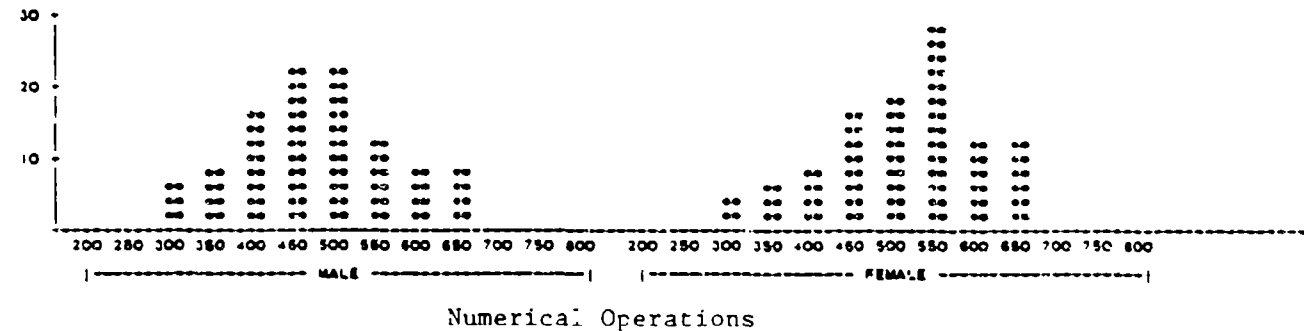
Percentage



Percentage



Percentage



Percentage

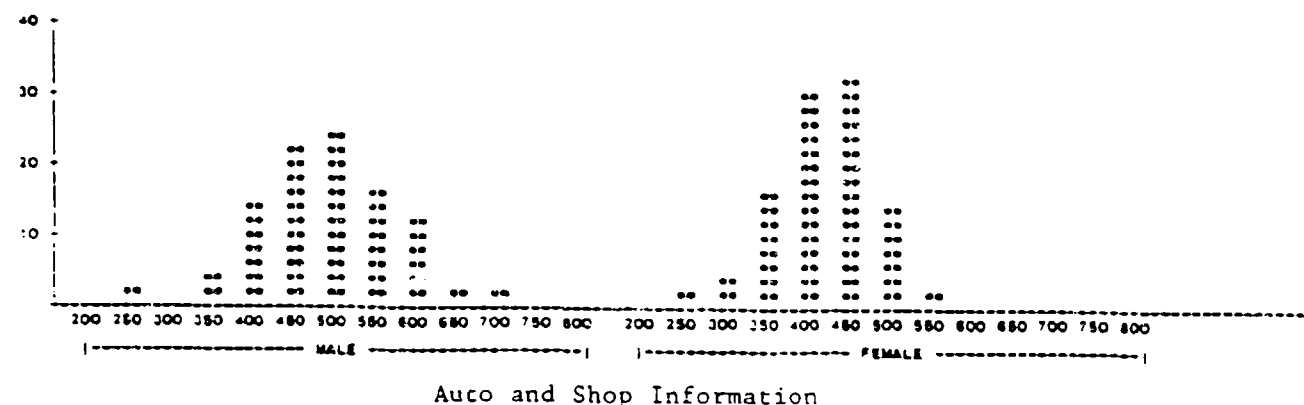




Figure 4.2-7 Hispanics with some high school by sex.

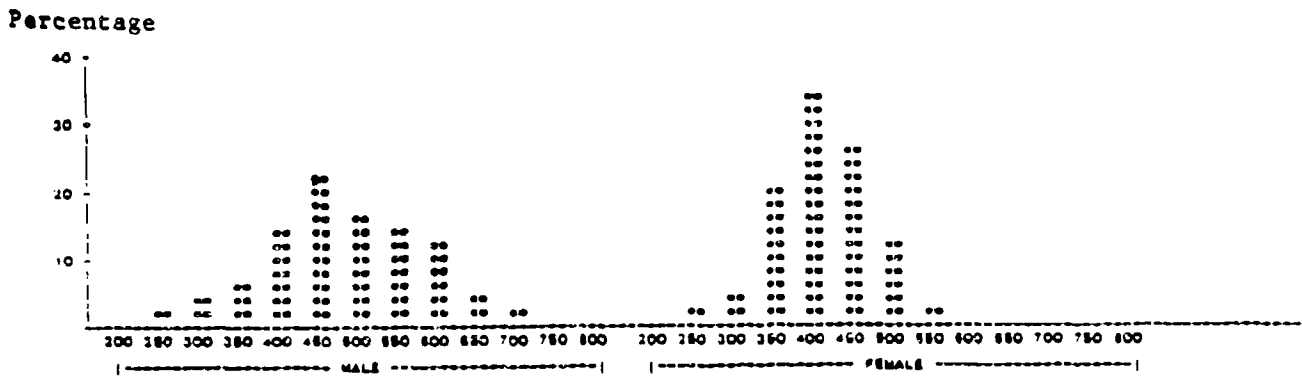
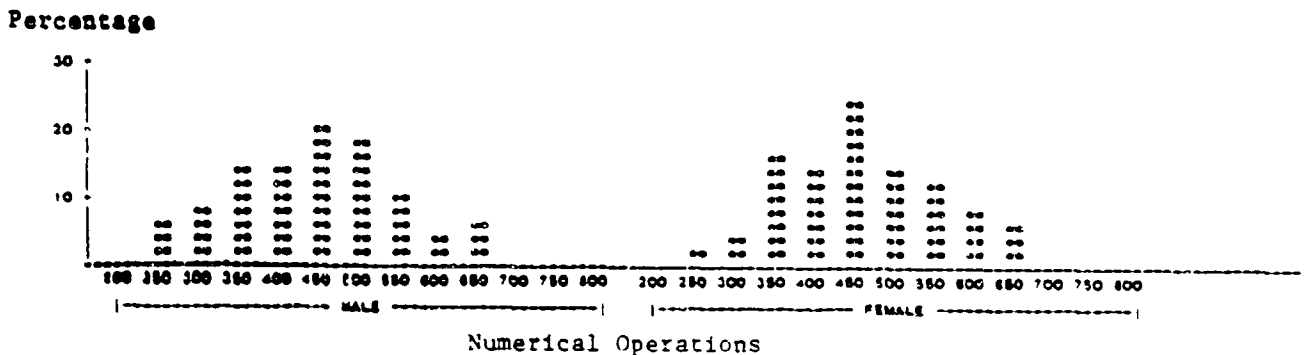
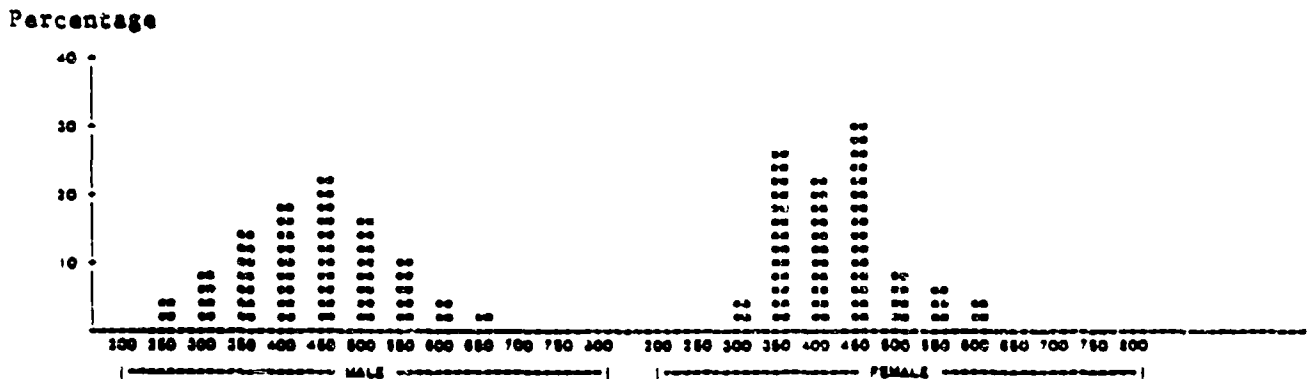
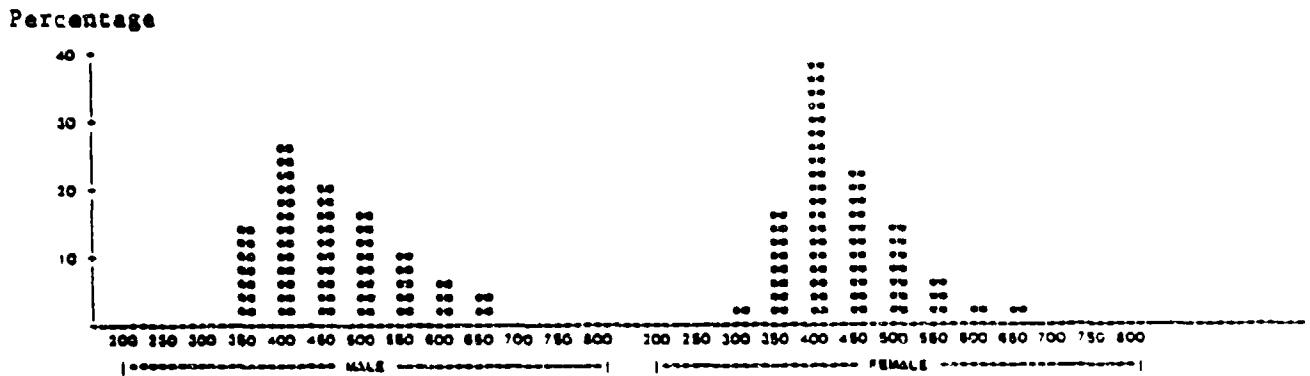
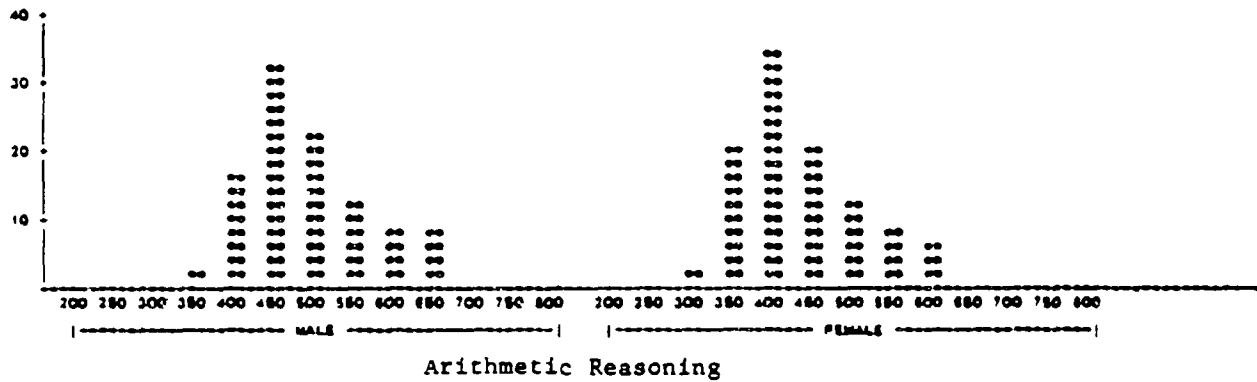
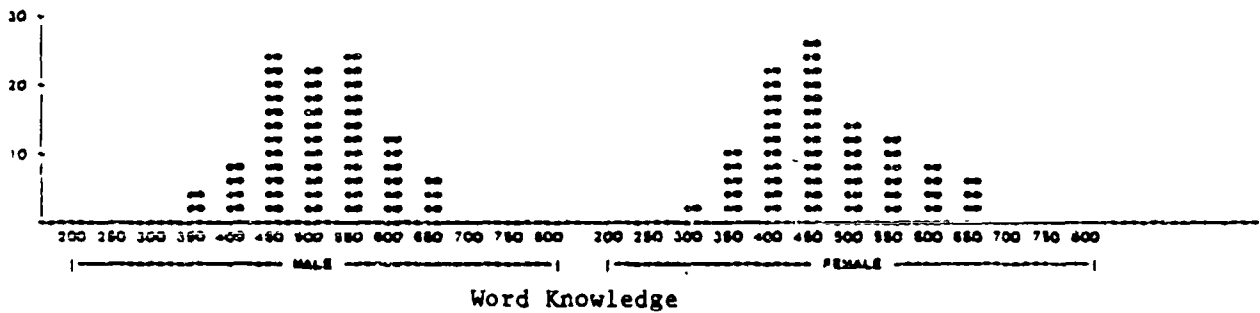


Figure 4.2-8 Hispanics with high school by sex.

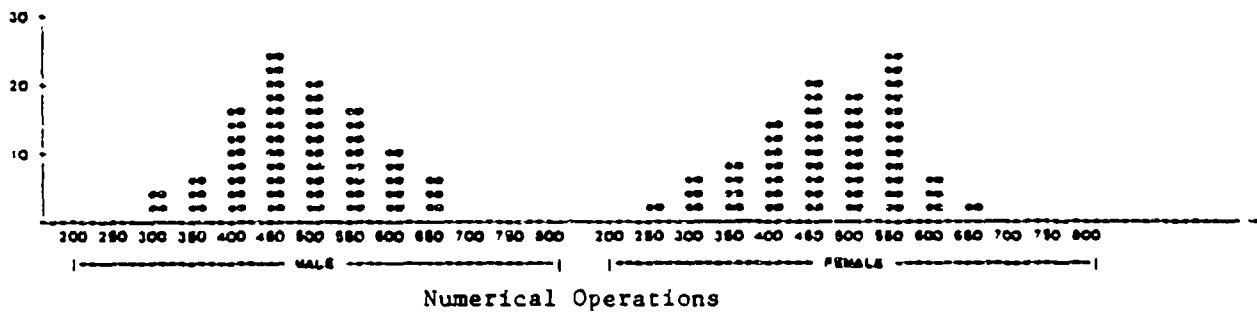
Percentage



Percentage



Percentage



Percentage

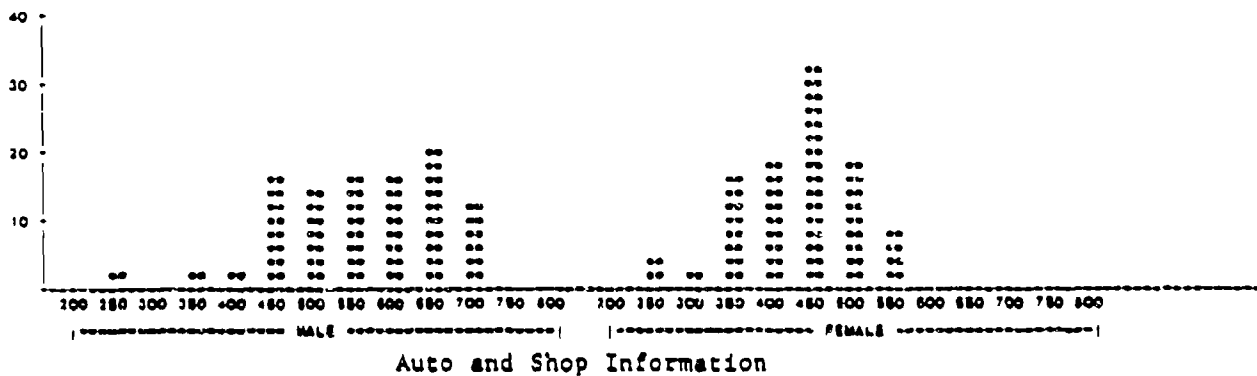
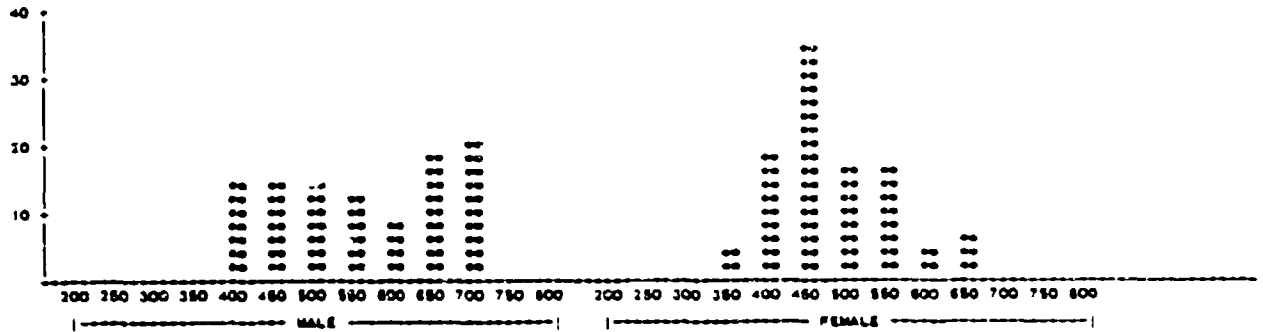


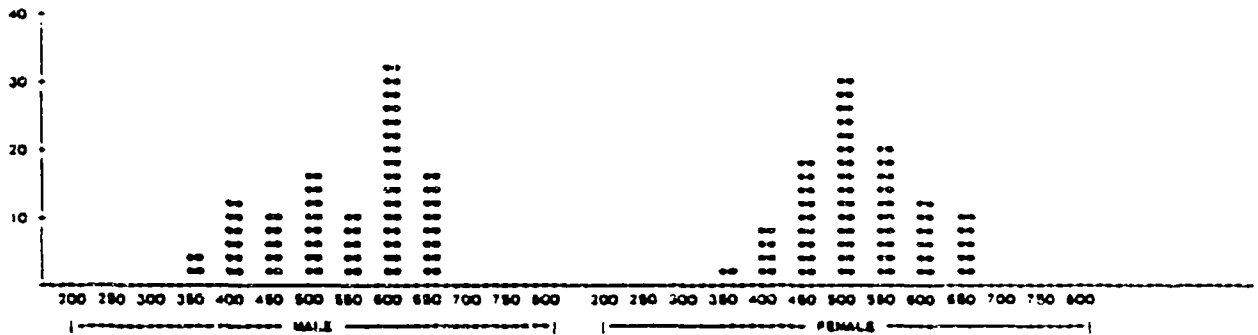
Figure 4.2-9 Hispanics with some college by sex (includes college graduates).

Percentage



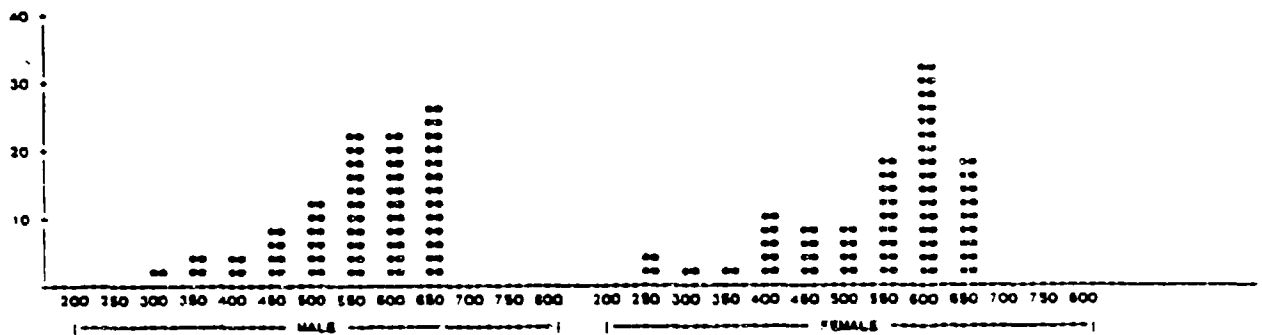
Arithmetic Reasoning

Percentage



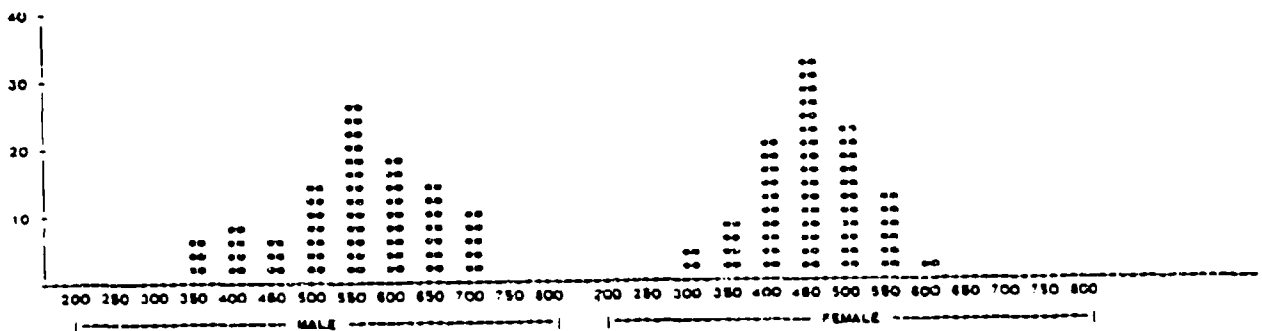
Word Knowledge

Percentage



Numerical Operations

Percentage



Auto and Shop Information

have completed some high school, those who have graduated from high school but have not attended college, and those who have attended or graduated from college. It is in the comparison of these groups that the effect of education on the shape of the distributions is most apparent.

The ASVAB tests are aimed primarily at those who have completed high school but not attended college, the group which is at present the main source of recruits for the Armed Services. The items in the tests have been chosen so as to obtain good discriminability among subjects in this ability range. We see from the histograms in Figure 4.2-2 that among White male and female subjects the Arithmetic Reasoning scores fulfill this condition well and have unimodal and essentially symmetric distributions over the range. The Word Knowledge and Numerical Operations tests are somewhat easier for these subjects, however, and there is some piling up of cases on the highest score range, 650-700. A certain number of subjects appear to be reaching the test ceiling, that is, answering all items correctly, and thus obtaining the highest possible scale score. This produces a slight skew in the distributions toward the lower ranges where, because the tests have adequate numbers of easy items, the scores are not restricted. In other words, these tests show very little "floor" effect.

The Auto and Shop Information distributions show at all education levels the marked sex differences that are noted and discussed in Chapter 7.

Among Whites with some college education (Figure 4.2-3), a larger proportion of the sample is reaching the test ceiling and the skew of the score distributions is much more apparent. This is especially true of the rather easy Word Knowledge and Numerical Operations tests. Only in the Arithmetic Reasoning test scores of the female subjects, for whom the arithmetic test is more difficult than for males (see Chapter 7), is the score distribution nearly symmetric at the college level.

For the Arithmetic Reasoning test especially, the Blacks at the high school level (Figure 4.2-5) find this test more difficult to an extent that the score distributions are skewed in the opposite direction. The majority of scores fall below the general population mean of 500, and only a small proportion of these subjects are found in the high score ranges. This effect is somewhat less apparent in Word Knowledge and hardly evident in Numerical Operations. The latter test shows an essentially symmetric distribution in the Black high-school-only education group.

Among Blacks (Figures 4.2-4, 5, and 6) another phenomenon appears. There is an indication of bimodality in the score distributions, with a small group of Blacks scoring at levels comparable to the Whites, while a larger group of Blacks cluster at lower score levels. This effect is apparent in the Numerical Operations scores of Black females with some or completed high school education and in the Arithmetic Reasoning scores of Black males at these education levels.

In the Word Knowledge scores of Blacks, one sees a more general shifting upward of the distributions combined with a larger proportion of subjects in the highest score ranges. Because these tests are rather easy, any tendency towards bimodality of score distribution that might exist in the population is obscured by test ceiling effects.

The fact that no similar evidence of bimodality appears in the Auto and Shop Information scores of Blacks suggests an effect of education may be involved. The small proportion of Blacks who score in the range of the highest ranking Whites must be those who are participating in and meeting the standards of the educational programs where Whites are the dominant majority. These would include the college preparatory programs in high school and the more demanding courses of study in college. Some of the effects of the

different educational emphases and experiences between Blacks and Whites are discussed in Chapter 5.

The results for the Hispanics (Figures 4.2-7, 8, and 9) are similar but more complex. For the Arithmetic Reasoning test, Hispanic male subjects show similar evidence of bimodality at the college level. The score distributions for high school and college differ considerably, with a substantial proportion of Hispanic college students falling in the high ranges of the Arithmetic Reasoning scores. The Hispanic females, on the other hand, show only a slight shift of the score distribution toward higher values, between high school and college, and only a very small excess of females appears in the highest score category. This is consistent with the impression gained elsewhere in the data analysis (see Chapter 7), that relatively few female Hispanic students pursue quantitative programs in college.

Although there is no evidence of bimodality of the Word Knowledge and Numerical Operations scores for Hispanics, Hispanic males participating in college programs are much more heavily represented at the high score levels. It appears that a somewhat larger proportion of well-prepared male students is represented among Hispanics who go to college than is true of the female students. Other evidence for this conclusion will be discussed in Chapters 5 and 7.

The tendency of the scores to depart from normal distributions within subclasses of the demographic design is, of course, much more exaggerated in this broad classification by sex, sociocultural group and education than is the case for the six-way design including socioeconomic class, region, and mother's education. Regrettably, the numbers of cases in the subclasses of the six-way design are too small to allow an empirical assessment of the shapes of distributions. We will proceed on the assumption that the

additional classes largely remove the evidence of bimodality seen here, thus bringing the data much better into agreement with the assumption of similar normal distributions within subclasses. If one wanted to answer the question of whether, for example, Blacks pursuing college programs are more heterogeneous in their preparation than are comparable Whites, a different allocation of sampling resources would be required. That is, the study would have to include many more Black subjects at the college level. The same would be true with respect to possible sex differences in the preparation and attainments of college bound Hispanics. With the present data, we must confine ourselves to the analysis and interpretation of the broader aspects of the demographic design.

One effect of the skewing of the score distributions in the high ability groups is that the variances of the measures are not entirely homogeneous. They tend to be smaller in the high ability groups. Because the numbers of subjects in these groups are relatively small, however, the lack of homogeneity has little effect on the pooled estimate of the within-cell variance. The estimated variances and covariances of the test scores depend much more heavily on the data from the large groups of subjects who have 9-11 years of education or are only high school graduates. For these groups the assumptions of normally distributed test scores within groups and homogeneity of variance more closely obtain. For this reason, and in view of the well-known robustness of analysis of variance techniques to departures from within-cell normality and homoscedasticity, we consider the ASVAB scores analyzed in this study to satisfy sufficiently well the assumptions of the multivariate analysis of variance. Most of the effects observed in this study are so large and clearly significant that only gross failures of assumptions could obscure them. In some of the more marginal two-factor and three-factor interactions,

however, the conclusions may be more sensitive to the assumptions on which they are made. In these cases, we fall back on the plausibility and interpretability of the result to convince us of its reality. The points at which some doubts about the significance of results may arise are few in comparison with the many clear-cut findings reported in Chapters 5, 6, and 7.

#### 4.3 Tests of the Model

In the analysis of variance of unbalanced designs, all of the estimated effects are in general intercorrelated and the actual numerical values obtained are influenced by the number of terms included in the model. Following the general experience that high-order interactions are rarely seen in nature, we attempt to account for the data by a model that includes only main-class effects and low-order interactions. According to the hierarchical principle, all lower-order effects represented in a higher-order interaction must be included in the model. For this reason, we test the partial effect of adding higher-order interaction terms after the variation between subclass means due to lower-order effects and interactions has been accounted for. Our objective is to exclude from the model those interactions that contribute nothing to between-subgroup effects.

The results of the multivariate analysis of variance for choosing terms in the model are shown in Tables 4.3-1 and 4.3-2. In each case we show the multivariate test statistic for the hypothesis that the corresponding effect or interaction is null. They are presented in the form of Rao's F-approximation for the likelihood ratio statistic. These F-statistics may be compared with the .05 and .01 points of the F-distribution, for the approximate number of degrees of freedom, shown in the following two lines. In addition, we show for each of the ASVAB tests the univariate F-statistic on the same null hypothesis, and at the bottom of the table we give the .05 and .01 points with which they may be compared.



TABLE 4.3-1  
MULTIVARIATE AND UNIVARIATE TEST STATISTICS FOR EFFECTS OF BACKGROUND FACTORS: FIRST ANALYSIS

	Subjects Education (HCC)	Economic Status (ECS)	Sociocultural Group (SCG)	Sex	Mother's Education (HCC)	Age (Age)	HCC x SCG	SCG x SCG	SCG x HCC	Age x Lag	SCG x Lag	SCG x HCC	HCC x ECS	SCG x HCC x ECS	SCG x ECS x Lag	Residual						
Likelihood ratio:																						
P-approximation																						
F (.05)	208.1	190.8	210.7	896.3	31.4	16.2	5.0	24.3	9.0	3.3	2.9	0.7	3.6	1.9	0.9	1.7	1.0	1.1	1.04			
F (.01)	1.5	1.0	1.5	1.8	1.4	1.5	1.3	1.6	1.5	1.5	1.0	1.3	1.5	1.4	1.3	1.3	1.6	1.3	1.00			
F (.01)	1.7	2.3	1.7	2.3	1.6	1.7	1.5	1.9	1.7	1.7	1.7	2.3	1.5	1.6	1.4	1.6	1.5	1.9	1.5	1.00		
Univariate F																						
UK	1679.8	1259.2	1222.2	0.6	216.2	25.2	7.1	2.4	4.1	1.0	11.0	2.2	6.2	2.6	0.6	0.8	1.7	0.5	3.1	3.4	2.5	1.24
PC	1153.2	957.9	800.8	80.8	147.5	7.2	6.2	1.2	6.4	2.3	16.0	1.5	7.5	1.4	0.6	1.3	0.9	0.2	4.1	3.5	2.1	1.17
AR	1018.9	920.0	1042.4	296.4	143.5	14.1	19.5	10.6	15.0	20.4	6.9	0.2	8.0	1.9	1.0	2.4	3.6	0.9	2.7	3.5	1.1	1.10
MO	859.0	685.0	501.4	143.9	87.1	23.6	2.6	0.2	7.1	1.4	0.5	4.3	5.3	1.4	0.5	0.7	2.0	0.8	0.8	5.4	2.0	1.23
CS	953.9	550.2	586.2	590.2	55.1	17.2	3.9	3.3	4.3	0.3	7.1	6.2	3.5	0.9	0.3	1.2	1.2	0.8	1.1	0.2	0.9	1.07
CS	1139.3	1054.5	1236.4	596.6	198.5	4.0	4.5	24.5	15.9	1.3	10.4	1.1	4.9	3.1	0.8	2.2	3.5	1.1	3.6	4.6	1.7	1.12
HK	1013.7	1094.7	585.9	33.3	205.2	17.6	11.3	4.1	19.3	6.7	12.5	0.6	6.2	2.1	2.1	0.3	3.3	1.0	4.8	6.9	0.8	1.23
MC	595.3	974.4	1286.9	2061.	98.9	22.2	10.8	78.0	21.1	9.2	6.9	10.4	7.7	2.6	0.5	3.8	2.9	0.9	3.4	2.3	0.7	1.02
LI	968.0	976.1	1251.6	2153.9	91.4	8.3	11.6	36.9	34.9	2.3	3.5	12.8	4.3	5.3	0.1	4.3	7.9	0.3	1.5	4.2	3.5	1.12
AS	614.2	1094.7	1461.5	4926.3	31.6	52.2	6.6	154.2	38.8	1.8	2.2	40.0	6.6	6.1	0.5	1.4	4.3	0.4	0.9	1.1	2.6	1.09
F (.05)	2.6	3.8	2.6	3.8	2.4	2.6	2.1	3.0	2.6	2.6	2.6	3.8	2.1	2.6	2.6	2.4	1.9	2.4	2.1	3.0	2.1	1.00
F (.01)	3.8	6.6	3.8	6.6	3.3	3.8	2.8	4.6	3.8	3.8	3.8	6.6	2.8	3.8	3.8	3.3	2.5	3.3	2.8	4.6	2.8	1.00

TABLE 4.3-2

MULTIVARIATE AND UNIVARIATE TEST STATISTICS FOR EFFECTS  
OF BACKGROUND FACTORS: SECOND ANALYSIS

	HGC through Sex x Region	Sex x MHGC	ScG x MHGC	ScG x HGC x EcS	Sex x ScG x EcS	Residual
Likelihood ratio: F-approximation		2.0	2.1	1.6	1.8	1.04
F (.05)		1.4	1.3	1.3	1.6	1.00
F (.01)		1.6	1.4	1.5	1.9	1.00
WK	(See Table 4.3-1)	0.8	1.8	3.1	3.3	1.24
PC		1.3	1.0	4.0	3.6	1.17
AR		2.4	3.7	2.6	3.6	1.10
NO		0.7	2.0	0.8	5.4	1.23
CS		1.2	1.2	1.1	0.2	1.07
GS		2.2	3.5	3.6	4.5	1.13
MK		0.3	3.2	4.4	7.0	1.23
MC		3.9	2.9	3.6	2.1	1.02
EI		4.3	2.9	1.6	4.2	1.14
AS		1.4	4.3	0.9	1.1	1.10
F (.05)		2.4	1.9	2.1	3.0	1.0
F (.01)		3.3	2.5	2.8	4.6	1.0

As is apparent in Table 4.3-1, these statistics are extremely large for the main effects and for such two-factor interactions as sociocultural group by highest grade completed, sex by sociocultural group, sex by highest grade completed, sociocultural group by economic status, highest grade completed by economic status, sex by economic status, and sociocultural group by region. At this point in the model building, significance levels drop off sharply. There is some evidence of sex by region effects, especially for the male-specialized subjects of Auto and Shop Information and Electronics Information; and there is no evidence of interaction of region by socioeconomic status. There appear to be very marginal interactions between sex and mother's education, and between sociocultural group and mother's education. As mentioned above, the latter have a clear interpretation and probably are real. We are less certain about sex by mother's education. There is no evidence whatsoever of interaction between mother's education and economic status, so this term can be dropped from the model.

Among the triple interactions we consider only sociocultural group by highest grade completed by economic status, and sex by sociocultural group by economic status. The latter is of borderline multivariate significance and possibly could have been omitted. However, it has a fair-sized effect for the Mathematics Knowledge test, which is of special interest in the discussion of sex differences (Chapter 7). For this reason sex by sociocultural group by economic status was retained in the model. Both of these interactions are highly specific to certain of the groups and certain of the variables. In the case of sociocultural group by highest grade completed by economic status, the interaction, as we shall see in Chapter 5, is entirely due to the lack of any economic status effect in Whites with one or more years of college. In the case of sex by sociocultural group by economic status, the effect is confined to female Blacks, primarily in the Mathematics Knowledge test.

The final interaction that is considered separately--sociocultural group by economic status by region--is clearly nonsignificant and can be dropped from the model. All remaining interactions and the remaining triple interactions that have little plausibility are pooled and tested jointly against the within-cell error. Although, because of the enormous number of degrees of freedom in this test, the multivariate test statistic for the residual is significant and some of the univariate test statistics are significant, the F-ratios for the individual measures cluster so closely around their theoretical value of 1.0 that there seems little reason to attempt to include any additional terms in the model for estimation. All subsequent analysis of the data is therefore based on the model with 69 degrees of freedom, including all of the terms in Table 4.3-1 except the sociocultural group by economic status by region interaction, the mother's education by economic status interaction, and the region by economic status interaction. Statistics for the analysis under the abbreviated model are shown in Table 4.3-2.

#### 4.4 Interaction Displays

The results of greatest descriptive interest are the estimates of the significant two- and three-factor interaction effects, and the main effect of mother's education, which shows little interaction with other factors. These estimates are displayed in subsequent chapters in the form of interaction plots of the type illustrated in Winer's (1971) book on analysis of variance. The figures in these plots are formed by adding together the effects that contribute to each subclass. This includes, of course, general mean and main-class effects, plus the particular interaction terms represented by the classes that jointly define the cell. As a further aid to interpretation of these plots, each point is bracketed by a standard error bar (plus and minus

one standard error), computed from the variances and covariances of the estimated effects that are summed to obtain the subclass values. The variance and covariance factors required in these calculations are given in the extended printout of the MULTIVARIANCE program. To be conservative, we have allowed for the effect of intraclass correlations on the estimated error variance by multiplying the nominal standard error by 1.4. This corresponds to a design effect of 2, which appears to be a definite upper limit in these data.

Especially in the plot of the sociocultural group by highest grade completed by economic status interaction, the standard errors vary greatly. This is a reflection of the widely differing numbers of subjects in these cells of the design. There are very few cases with only 0-8 years of education in some groups and few with 13+ years in others. For most cells, however, the sample size is large enough to locate the interaction line within a narrow range. All of the standard error bars, long or short, have a two-thirds chance of including the true line. The actual line is drawn at its most likely point--namely, the middle of the two standard error intervals.

It is emphasized that these interaction plots are not the population marginal means for the corresponding subclasses. Population marginal means of complex cross-classifications do not in general have a useful interpretation when there are significant effects and interactions with the dimensions of the table that are collapsed when obtaining these means. The effects of the collapsed ways of classification are mingled in the marginal means in a way that makes interpretation difficult, if not impossible. The interaction plots depict only those between-group effects that occur partially when the effects of the other ways of classification are eliminated. For example, the plots of sociocultural group by highest grade completed by economic status are free of the effects of region and of the interaction of region with other ways of

classification. The plot represents the best estimate of the common socio-cultural group by education by economic status interaction within regions.

This is not to deny that in surveys such as the Profile study the weighted marginal averages have a practical utility. Distributions in the margins are essential in calculating classification rates for the marginal classes. For example, the weighted marginal means of the two sexes and the sociocultural groups are shown in Table 4.3-3. The figures in Table 4.3-3 have, for example, the following implications:

Suppose a score of 500 or better on Auto and Shop Information was required for entering Vehicle Maintenance School. Assuming a within-marginal-class standard deviation of 100 (a slight overestimate in this case), the percentage of men in the general population represented by the Profile sample who would qualify is 70.2; the percentage of women would be 30.9. Similarly, the percentage of Whites, Blacks, and Hispanics qualifying would be 57.5, 20.1, and 30.9, respectively. Such figures are important in the practical use of the tests for personnel selection, but they do not bear as directly on the question of sources of differences as do the estimated effects in the interaction plots.

In the case of the complex interaction plots, the source of the interaction may be confined to certain variables and to certain subgroups. Some indication of the significant variable can be obtained from the individual F-tests shown in Table 4.3-1 and 4.3-2. As a convenience to the reader, tests for which the depicted interaction is clearly significant are marked with an asterisk in the various displays in Chapters 5, 6, and 7. In many cases, however, direct inspection of the plot with a view to the possible sources of the interaction in psychological or sociological terms makes clear whether and why the composite null hypothesis is rejected.

TABLE 4.3-3  
WEIGHTED MARGINAL MEANS OF THE ASVAB STANDARDIZED  
SCORES FOR MEN AND WOMEN AND FOR WHITES,  
BLACKS, AND HISPANICS

Test	Men	Women	White	Black	Hispanic
Word Knowledge .....	501	504	522	411	445
Paragraph Comprehension .....	493	512	520	423	449
Arithmetic Reasoning .....	515	489	520	420	449
Numerical Operations .....	490	514	517	430	462
Coding Speed .....	481	524	516	430	473
General Science .....	521	483	521	415	445
Mathematics Knowledge .....	506	498	517	434	451
Mechanical Comprehension ....	538	465	520	415	452
Electronics Information .....	537	465	520	417	445
Auto and Shop Information ...	553	450	519	416	450

## CHAPTER 5

### SOCIOCULTURAL EFFECTS

The eight-year age cohort of young people in this study is of special interest because it is among the first to benefit from legislation and programs of the 1960s and 1970s aimed at extending educational opportunities. Early in the 1960s, it could no longer be ignored that some segments of the society were experiencing unmitigated intergenerational poverty at a time when most Americans were enjoying prosperity and upward mobility in an expanding economy. While the historical forces of segregation and social isolation based on race and national origin had undoubtedly contributed to such poverty, legislative and judicial interdiction of these practices was clearly not sufficient to improve conditions. The majority of individuals in these excluded groups were still not acquiring the skills and credentials required in an increasingly technical society. Children of the poor and socially isolated were still showing levels of educational attainment and scholastic achievement well below those of their more advantaged age mates. Public education was not succeeding as the great "equalizer" that it was supposed to be. Black, Chicano, Puerto Rican, and other minority status children were not developing skills at the level necessary to compete for jobs and social position.

Despite the general agreement that restricted socioeconomic mobility was strongly related to deficient educational achievement, there was no consensus on the factors contributing to school failure, how they were mediated by social class and ethnic group membership, or what remedial action should be taken. Numerous intervention programs, varying in focus and content according to the theoretical orientation of their proponents, were formulated and put



into practice. Supported by federal funds allocated for the War on Poverty, Project Head Start was launched on the national level for the purpose of attenuating the sociocultural handicaps that deprive children between the ages of 3 and 5 of the potential benefits of public education. Increased potential for upward economic and social mobility was to be the result.

At the same time, the school desegregation movement was gaining momentum. To many, school segregation was seen as a major contributing factor to the achievement differences observed between White and non-White students. The National Association for the Advancement of Colored People (NAACP) initially argued that differences in the allocation of educational resources to Black and White schools resulted in inferior education for Blacks, and that equality in the distribution of school resources could efficiently be achieved through school desegregation. The Supreme Court agreed with this view, as indicated by the 1954 decision (*Brown vs. Board of Education*), and accelerated the desegregation movement. The expectation was that all-Black schools would eventually disappear, and Black students would have an unrestricted opportunity to interact with middle-class White peers who would initially serve as role models of success in school and in later life (Ogbu, 1978).

The Profile data provide an opportunity to assess the broad effects of this era of innovation on the vocational preparation of young people of some of these sociocultural groups. These data cannot, of course, be used directly to measure the effectiveness of various intervention programs or of school desegregation because there is no way to know whether the young people in this sample participated in such programs or attended desegregated schools. However, all respondents were either of school age, or subsequently achieved school age, during the height of this national emphasis on equal educational opportunity for all. Whether we may reasonably expect any

substantial effects from this movement on the relative achievement of the different sociocultural groups depends upon the extent to which reform, both educational and social, was actually translated from good intentions to effective action. The Profile data only allow us to view the situation as of the spring and summer of 1980.

#### 5.1 Sociocultural Group by Economic Status by Education Effects

In the sections that follow we examine the group differences seen in the ASVAB test scores. As might be expected, a pattern of increasing test scores with educational level is seen for all groups in all tests. Respondents in each group who had completed some college scored highest on the tests, while those with less than 9 years of education scored lowest. Although the performance of all categories of young people improves with education, for many of the tests significant differences in average performance appear between sociocultural and economic groups who have completed the same amount of formal education. Generally, Whites scored higher than Hispanics, who scored higher than Blacks, and the nonpoor in each of the groups scored higher than their poor counterparts. However, significant interactions were observed between the effects of sociocultural group, economic status and education.

These interactions are displayed in the plots in Figure 5.1a and b. The performance of poor and nonpoor Whites, Hispanics, and Blacks with 0 to 8, 9-11, 12, and 13 or more years of education is shown in terms of the standard scores of the total probability sample for each subtest (mean=500, standard deviation=100). The interpretation of these effects is simplified by the fact that they fall into the four groupings of the tests suggested by the within-group factor analysis--namely, specialized knowledge (Auto and Shop Information, Mechanical Comprehension, and Electronics Information), general verbal ability

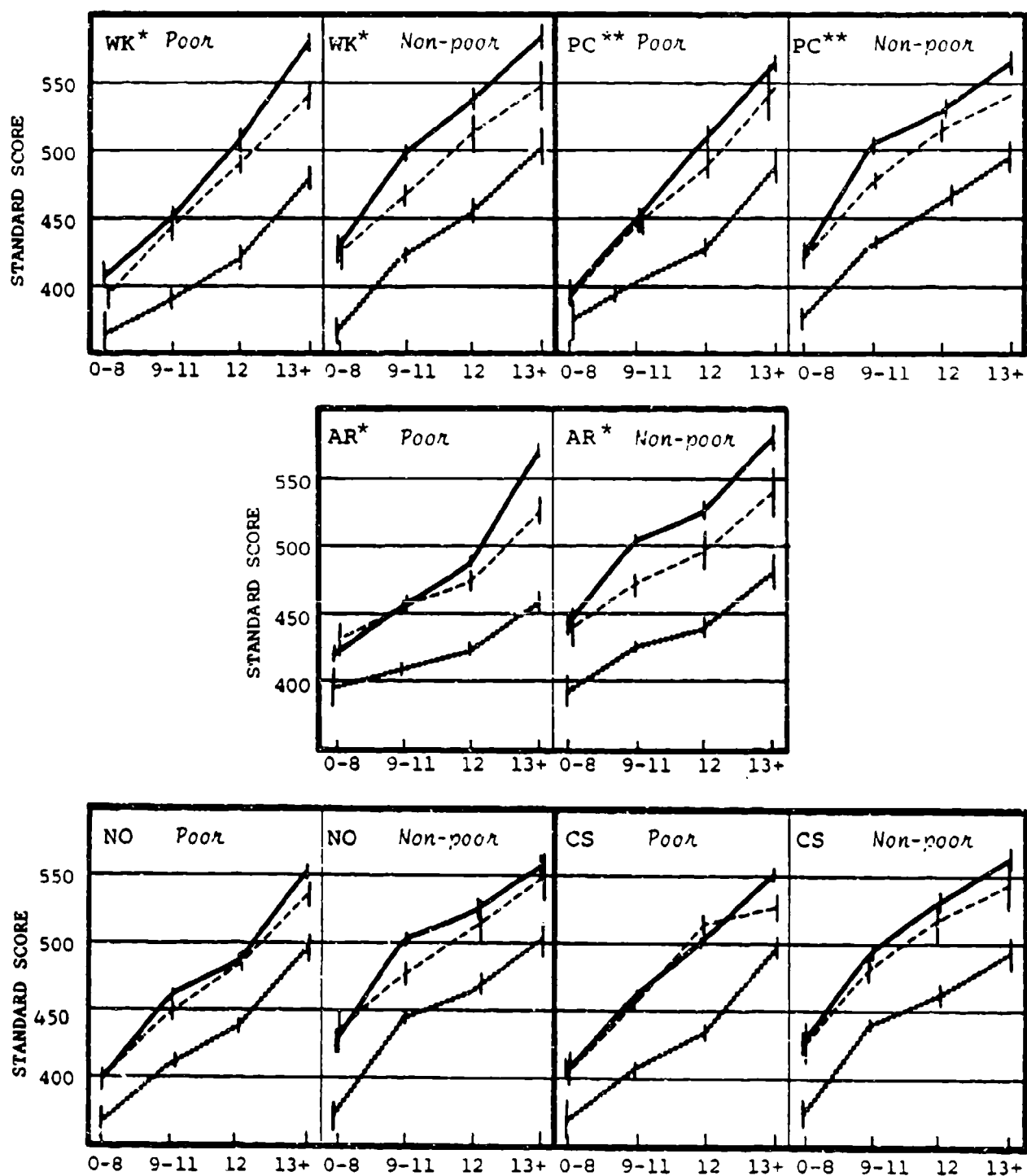


Figure 5.1a Interaction of sociocultural group, economic status, and education (highest grade completed)

White——— Hispanic----- Black.....

\*F > 1.25F(.05) \*\*F > 1.25F(.01)

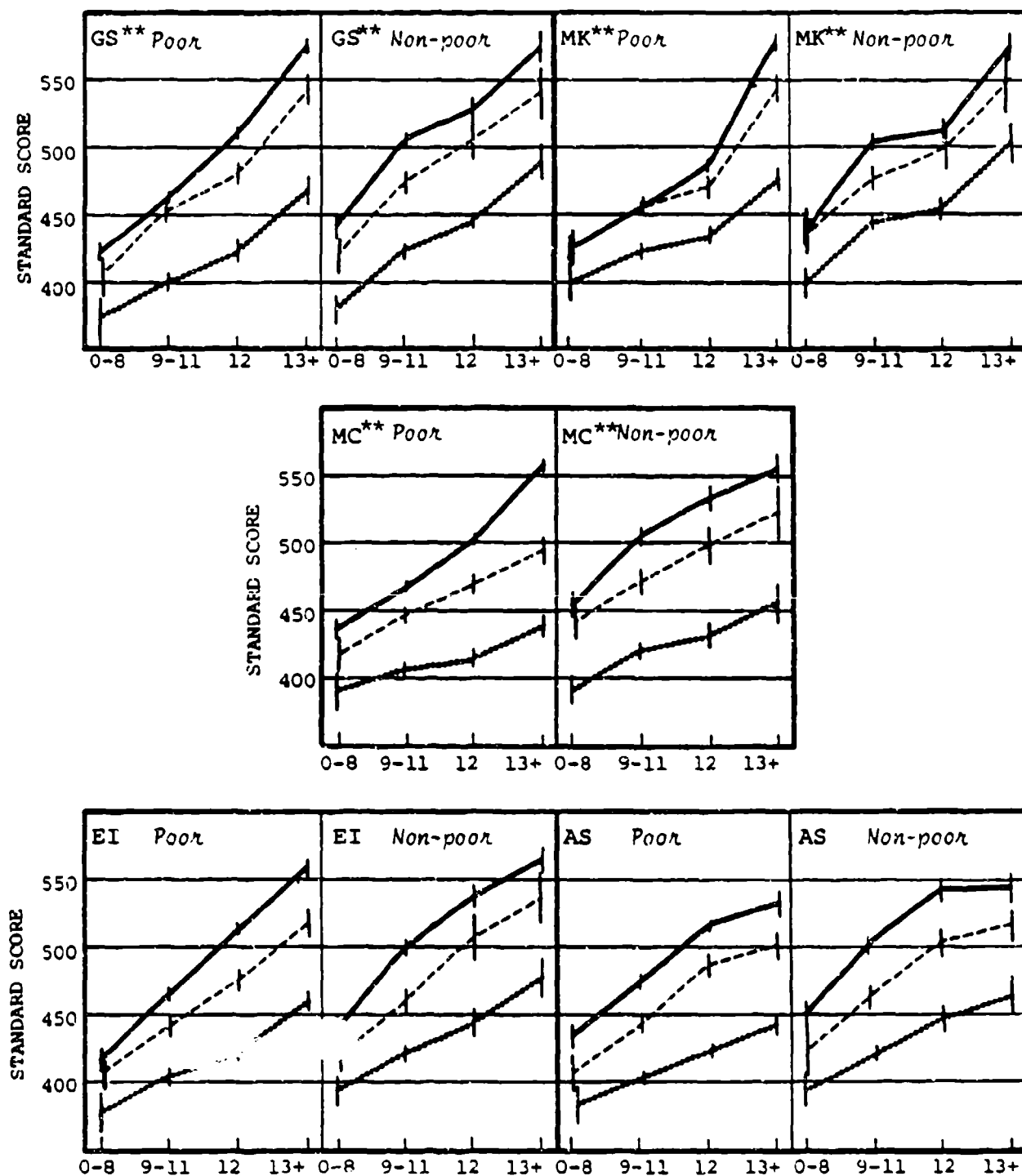


Figure 5.1b Interaction of sociocultural group, economic status, and education (highest grade completed)

White ——— Hispanic - - - - - Black ······

\*\*F > 1.25F(.01)

(Word Knowledge and Paragraph Comprehension), fluent production (Numerical Operations and Coding Speed), and quantitative ability (Arithmetic Reasoning and Mathematics Knowledge). The General Science test shares some of the features of both specialized knowledge and general verbal ability.

Specialized Knowledge. Mechanical Comprehension, Electronics Information, and, in this context, General Science, show very similar patterns of interaction. (The Auto and Shop subtest shows a different pattern of interaction and is discussed separately below.) It is obvious from the plots of these tests that all of the sociocultural groups show improved performance with increasing education. Through 12 years of education, economic status has a similar effect in all the sociocultural groups, amounting to about a 30 point advantage of nonpoor respondents over their poor counterparts. However, at the highest educational level (some college) an unexpected interaction of sociocultural group and economic status appears. In the specialized knowledge tests, and indeed in nearly all the tests, poor Whites equal and even exceed the performance of nonpoor Whites among respondents with some college education. A similar effect is much weaker in the Hispanic and Black subsamples. This interaction has a simple explanation.

That poor Whites at the highest educational level equal or outperform their higher-income counterparts is undoubtedly the result of using the OMB definition, discussed in Section 2.4, to classify college students as below the poverty line when they are living on low incomes independently of their parents. Because many young people in college today declare themselves as independent of their parents, a considerable number of White college students classified as poor in this study are, in fact, probably from more economically advantaged backgrounds. As a result, the performance of White respondents with some college is largely independent of their income classification.

Low-income Whites with 12 or less years of education reflect more accurately their actual economic class membership, and are about 25 points below nonpoor Whites on most tests.

In the Black subsample the effect of income-status on the performance of respondents with some college is different from that of Whites and, to a somewhat lesser extent, from that of Hispanics. For Blacks, this effect is more similar to that seen at the lower educational levels, which suggests that the income classification of Black college students is a more accurate index of their economic class membership than it is for Whites. Perhaps this is due to relatively more Black college students living with their parents, or to those living independently not having incomes greatly different from those of their parents.

Comparing nonpoor Whites and nonpoor Hispanics, we find very similar effects of education on the specialized knowledge tests. In General Science, these subsamples show an average difference favoring Whites of about 25 points, and in Electronics Information about 30 points. In Mechanical Comprehension, there is very little difference at the lowest educational level (0 to 8 years). At the 9-11 educational level and above, however, Hispanics are about 35 points lower than Whites.

Blacks respond somewhat differently to the specialized knowledge tests. The plots for nonpoor Blacks and nonpoor Whites and Hispanics show that in all tests except General Science, the difference in performance tends to increase with education through grade 12. (General Science is more like Word Knowledge in showing a nearly constant difference between Blacks and other groups.) In the still more technical Mechanical Comprehension and Electronics Information tests, nonpoor Blacks show considerably less gain in performance between the lowest educational level (0 to 8 years) and grade 12

than do the other groups. As a consequence, the 60 point difference in the scores of nonpoor Whites and Blacks seen at the lowest educational level increases to 100 points at the 12-year and the 13+ levels. This kind of effect would be expected if Blacks are receiving less exposure to this particular type of content in later years of high school. This would explain why, for Blacks, the gain due to completing high school is smaller for these tests than for more general tests such as Word Knowledge or even General Science.

The misclassification of self-reporting White or Hispanic college students as "poor," which we discussed above, makes the interpretation of the sociocultural group differences among subjects classified as poor more difficult. Because this effect is less apparent in Blacks, it makes the difference between Blacks and Whites or Hispanics artificially large at the college level. Differences between sociocultural groups at this level can be interpreted with confidence for the nonpoor subjects, but not for those classified as poor.

As we mentioned above, the Auto and Shop Information test, although certainly assessing specialized knowledge, shows a pattern of interaction with socioeconomic cultural groups, economic status, and education different from the other tests in this group. We believe that this pattern appears because auto and shop information is not primarily learned in school. As can be seen in Figure 5.1a, the Auto and Shop Information scores of respondents with some college are on average not much higher than those of respondents with 12 years of education. This is especially true of nonpoor Whites, who show no discernible difference in achievement between high school completion and some college. The trend is similar among nonpoor Hispanics, poor Whites and poor Hispanics, where the average respondent with some college scores only about 15

points higher than one who has only completed high school. In the Hispanic and White subsamples, performance in Auto and Shop Information is enhanced very little by higher formal education, unlike the effect for General Science, Mechanical Comprehension, and Electronics Information where formal study of scientific and technical subjects appears to improve performance considerably.

For Blacks in both economic classes, the influence of education on Auto and Shop Information is different from that of other sociocultural and economic groups. Especially among poor Blacks, performance increases steadily with increased educational level, while for the other groups, the difference in performance between 12 and 13+ years of education is relatively small. For Blacks, formal education beyond high school enhances performance in Auto and Shop Information about as much as the gain between some high school and completed high school. Presumably, more education for this group is associated with greater access to automobiles and machinery, probably through a connection between more education and greater economic resources, and this relationship does not apply to the same degree in the other groups. Relative to the Black community as a whole, Blacks who go to college are economically a more select group than are college-bound Whites, relative to the White community.

Interestingly, the Auto and Shop Information test is the only specialized knowledge test in which the difference in the average performance of poor and nonpoor Whites with some college favors higher-income Whites. The difference is small but may reflect a difference in access to automobiles during the college years between students living independently of their parents and those who are not.

At all but the highest educational level, nonpoor Hispanics equal and even exceed the performance of poor Whites in each of the verbal attainment



tests. Again, the fact that poor Whites show an advantage at the 13+ education level is largely an artifact of the above-mentioned misclassification of many economically disadvantaged White college students as poor. Nonpoor Hispanics also parallel the nonpoor Whites in their verbal attainment in English. The scores show an average difference favoring Whites, of about 30 points in Word Knowledge and about 25 points in Paragraph Comprehension at each educational level except the lowest. With 8 or fewer years of schooling the scores of nonpoor Hispanics and nonpoor Whites are the same. This indicates that higher-income Hispanics have relatively good English language facility, even though they lag somewhat behind their economic peers among Whites with some high school and with 13+ years of education.

Nonpoor Blacks tend to parallel the performance of nonpoor Whites and Hispanics in the verbal attainment tests. Unlike the trend of increasing differences with education between nonpoor Blacks and the other sociocultural groups seen for the specialized knowledge tests, in both Word Knowledge and Paragraph Comprehension there is a constant difference of about 75 points between nonpoor Blacks and nonpoor Whites and a difference of about 50 points between nonpoor Blacks and nonpoor Hispanics. In both the poor and nonpoor economic classes, the effects of education on the verbal tests are similar for the three sociocultural groups.

It is somewhat surprising to find that poor Hispanics equal or exceed nonpoor Blacks in English language facility at each educational level. In view of the literature that attributes poor Hispanics' scholastic achievement difficulties to language factors, this pattern of relative performance between these groups is unexpected. These data suggest, rather, that language factors may be more salient for the educational performance of Blacks than for Hispanics. We speculate later in this chapter on why the language proficiency differences between Hispanics and Blacks are seen in these data.

There is only a modest difference in the English language facility of poor Hispanics and poor Whites. In Paragraph Comprehension, their performance is identical at the lowest educational level and the 9-11 level. At the 12 year level and the highest level there is a constant difference between their average scores of about 20 points. In Word Knowledge, poor Whites show a slight advantage of about 15 points at the lowest educational level and the 12 year level. However, at the 9-11 year level, the scores of these two groups are nearly identical. For nonpoor Whites and Hispanics the effects of education on verbal attainment are quite similar.

The same is not true for poor Blacks. The difference in the verbal skills of lower income Blacks and their economic peers in the Hispanic and White subsamples increases with education through the 12th grade. This finding is consistent with that of Coleman et al. (1966), who reported that the gap in achievement between Blacks and Whites widened with continuing education. However, the present data indicate that this is the case only for poor Blacks, and only to the 12th grade. The effect of education on the verbal attainment of higher-income Blacks is similar to that seen in the other sociocultural groups.

Fluent Production. Since the Numerical Operations and Coding Speed tests involve minimal use of the English language, it might be expected that the performance of Hispanics would be nearly the same as that of Whites at the same economic level. This tends to be true, but more so for the poor Hispanics than for the nonpoor. Although nonpoor Hispanics and Whites at the lowest educational level perform identically in Numerical Operations, nonpoor Whites with some high school outperform nonpoor Hispanics by some 30 points on this test. In Coding Speed, a somewhat more modest difference of about 15 points in favor of Whites appears at this level. With high school completion

and some college, the difference between these groups shrinks slightly in Numerical Operations, but in Coding Speed there is a constant difference between some high school and some college.

The poor White and Hispanic groups show identical performance in Coding Speed across educational levels through the 12 year level. In fact, with 12 years of education, poor Hispanics slightly exceed poor Whites in Coding Speed. At the highest educational level, the poor Whites show an advantage of about 25 points, but this is again probably an artifact of the misidentification of a large number of White college students as poor. In Numerical Operations, the scores of poor Whites and Hispanics are essentially the same at each educational level from the lowest level through 12 years.

These data suggest that linguistic factors are not the only cultural variables affecting the attainment of Hispanics, especially nonpoor Hispanics. We will discuss later the work of Ramirez (1973), who suggests that culturally conditioned cognitive styles and motivational differences between Hispanics and Whites affect the relative attainment of these two groups.

Quantitative Attainment. Arithmetic Reasoning and Mathematics Knowledge scores show very similar patterns of education by sociocultural group by economic status interactions. The main source of the triple interaction is again the reduced economic status effect in Whites at 13+ years of education. This contrasts with about a one-fourth standard deviation difference on both tests between poor and nonpoor Blacks at this educational level.

Among nonpoor Whites and Blacks, the increased scores of those completing high school as opposed to those having 9-11 years of education is small relative to this increase for most of the other tests. This effect may have something to do with the concentration of general courses relevant to

Arithmetic Reasoning and Mathematics Knowledge in the first two years of high school. General math, algebra and geometry are taught to large numbers of high school students during these years, especially in middle-class communities. Many fewer students take the more advanced mathematics courses in the last two years of high school. For this reason, the difference in quantitative test performance from completing high school as opposed to having some high school is less than for the other tests that are less closely tied to this high school curriculum. Word Knowledge, for example, reflects more general intellectual attainment garnered from many sources both in and out of school. The level of such attainment reached by persons with increasing years of education increases steadily without any close relationship to the high school curriculum. The same steady increase is true of the other verbal ability tests and also perhaps of General Science. Those persons going on to college tend more often to take a full four years of mathematics, and will generally be exposed to it again in the first year of college. Thus, the increase in performance for the 13+ group is substantial for the quantitative tests. The data suggest that advanced mathematics education does not impact broadly the high school population, but rather tends to be concentrated on the college-bound.

For the technical knowledge tests such as Mechanical Comprehension and Electronics Information, the relevant high school courses are typically offered in the last two years of high school. The effect on these tests of completing four years of high school is more marked, and substantial score differences between educational levels are seen in Figure 5.1.b.

Minority Group Effects. Why Blacks show a lower benefit of education to specialized knowledge tests than to other tests is somewhat puzzling. The finding agrees, however, with other research showing that Black males generally express a lower level of vocational interest in scientific, technical,

and mechanical fields than do White males. Black males show relatively more interest in artistic, health and welfare, and business-clerical fields and in general show a more restricted range of occupational interests (Sewell and Martin, 1976; Williams and Whitney, 1978). As early as 1941 Witty et al. found that Blacks' vocational interests tend to be more "people oriented" while the interests for Whites are more "thing oriented". Bray (1972), who tested this early hypothesis by examining the majors of Black and White college freshmen, found that Blacks tend to major in the social sciences, education, and health fields, while White freshmen favor the biological and physical sciences. Attempting to account for these findings, Sewell and Martin (1976) suggest that the restricted range of vocational interests expressed by Blacks is the result of Black adolescents tending to aspire to occupations in which they have an opportunity to observe Black role models. Talented young Blacks choose careers elsewhere because they see few examples of successful Blacks in the scientific and technical occupations. Garfinkle (1975), in his examination of the occupations of Black workers between 1962 and 1974, confirms the paucity of Blacks in these "thing oriented" occupations relative to the more "people oriented" fields.

In contrast, Hispanics are proportionately more represented in scientific and technical fields than Blacks, which may explain the greater benefit of education to this group's performance on the specialized knowledge tests. Because young Hispanics in this country have more successful role models in the technical fields than Blacks, they may be expected to have more interest in such areas. There is evidence that more Hispanics than Blacks select courses related to these occupational areas: Sells (1980) reports that, in 1978, 25 percent of the Hispanic students in California were enrolled in mathematics courses leading to calculus, while 20 percent of the Blacks

were similarly enrolled. Moreover, Fernandez, et al. (1975) report from a sample of San Francisco high school students that about half of the black students in 10th grade were performing at less than the 6th grade level in mathematics, while only 17 percent of the Chicanos were performing below this level. Thirty-three percent of the Chicanos compared to only 18 percent of the Blacks received high grades in mathematics.

In a related statistic, Wilson and Portes (1980) report that in the Miami area, Cubans own 40 percent of the construction business and employ large numbers of their community in these businesses. Role models for Hispanics in the skilled trades are not lacking in the Southeast.

Sociocultural Group and Economic Status Differences in Relation to High School Attendance. Paragraph Comprehension, Arithmetic Reasoning, General Science, Mathematics Knowledge, and to a lesser extent Numerical Operations and Word Knowledge all show a curious form of interaction contrasting nonpoor Whites and Blacks with the other groups. It is especially apparent in the school-related context of Arithmetic Reasoning, Numerical Operations, General Science and Mathematics Knowledge. For these tests, the difference between subjects with 9-11 years of education and those with only 12 years is smaller for the higher-income Whites and Blacks than for the other groups. The result is a perceptible "step" in the middle of the curves for those groups. No such step is seen in Coding Speed, Auto and Shop Information or Electronics Information: it appears only to a lesser extent in Mechanical Comprehension. For the school-related tests, nonpoor Whites and Blacks with some college also show a relatively sharp increase in performance over those with no more than a high school education.

A possible explanation for this phenomenon is the effect of selective dropping out of high school. If less able respondents in the Hispanic groups

and the poor White and Black groups are more likely to drop out of high school before graduation, then the groups that retain most of their members through the 12th grade will be less select and score relatively lower on the tests. Current estimates indicate that the school dropout rate of Hispanics is higher than that of Blacks. The dropout hypothesis may explain some of the differences in the achievement patterns of Hispanics and Blacks. It does not, however, explain the difference in achievement between the Black and the White groups.

Contributing to the latter may be the effects of ability grouping or school tracking. Researchers report that a disproportionate number of Black (and other minority status) children are placed in low ability groups in school because of schools' reliance on general attainment tests to determine ability to learn (Brown, Carter, and Harris, 1978). Because minority status children generally do not perform as well on such tests as do majority culture children, they are placed in low ability groups where standards of achievement are low. In fact, ability grouping begins even at age levels where learning-readiness test scores are not available. Rist (1973) observes ability grouping as early as kindergarten before school personnel and teachers have any valid basis for determining ability to learn. Grouping in these early years appears to be strongly determined by children's physical appearance and verbal behavior. Similarly, Haskins, Walden and Ramey (1981) report considerable ability grouping in kindergarten and 1st grade classrooms based on teachers' informal observations of the children. These researchers concluded from classroom observations that the separation of lower and higher ability students may actually defeat the purposes and logic of intraclassroom ability grouping. They point to the adverse effect that ability grouping has on the type of instruction low-ability children receive. Moreover, early ability

grouping tends to establish a pattern of reduced expectations and cumulative educational deprivation for minority children, the pervasive influence of which is dramatically demonstrated by the work of Pedersen, Faucher, and Eaton (1978). These researchers present longitudinal data showing that teachers' expectations and attitudes toward children are correlated with the growth of children's verbal facility and the attainment of self-concept. These practices impinge on all children who are not identified with the majority culture, but their impact appears greatest on Blacks.

Linguistic Factors in Hispanics' Achievement. Because Spanish speakers are the largest linguistic minority in the United States, further consideration of linguistic factors is of interest. We have seen in the Profile data that the performance of nonpoor Hispanics in the nonlanguage tasks--Coding Speed and Numerical Operations--is somewhat lower than that of the comparable White group. This suggests that their lower average performance on language tasks may be attributable not only to linguistic factors, but to other cultural factors as well. In fact, among all the Hispanic subjects who took the ASVAB tests, only 85 requested that the instructions for taking the test be presented in Spanish (see Section 2.3). Note that the performance of Hispanics on the Paragraph Comprehension test is not much different from their performance on the nonlinguistic Numerical Operations and Coding Speed tests.

The nonpoor Hispanics exceed somewhat the poor Whites in Word Knowledge at all education levels except 13+, where poor and nonpoor Whites are indistinguishable. Poor Hispanics exceed nonpoor Blacks in this area and are more similar to poor Whites except at 13+ years of education.

The higher performance of the Hispanic groups relative to nonpoor Blacks at each education level is somewhat surprising in view of the long-



standing assumption that language skills of Hispanics hinder their educational attainment. Actually, one might have predicted that the Hispanics would show lower performance than the Blacks. In 1953, Anastasi and Cordova observed significantly lower intelligence test scores for Puerto Rican children than majority test norms would indicate. These investigators concluded that while the low socioeconomic status of these children contributed to their lower scores, "a solution to the language problem would seem a necessary first step for the effective education of migrant Puerto Rican children" (Anastasi and Cordova, 1953, p. 17). The present data indicate that the same cannot be said of the Hispanics generally in 1980: the language problem does not loom as large as it did in 1953.

There has been continuing controversy among educators and researchers regarding the extent to which bilingualism affects cognitive performance (Garcia, 1980; Carter, 1970), or even what bilingualism means as a construct. Cummins (1979) theorizes that a balanced proficiency in two languages can be highly beneficial to the child's cognitive development and functioning. Trends in English development of Spanish-speaking children in kindergarten and 1st grade observed by Garcia (1980) in a national study of English/Spanish bilingualism suggest that Spanish-speaking children show a decline in Spanish-language ability and a gain in English-language ability by 1st grade. The Spanish-language ability loss is thought to result from the monolingual educational curriculum. Fernandez, et al. (1975) report that Chicano high school students believe that learning English is very important for school performance and vocational prospects.

Portes, McLeod, and Parker (1978) found that the educational, occupational, and income aspirations of immigrant Mexicans are more strongly determined by their knowledge of English than is the case for immigrant

Cubans. These investigators surmise that familiarity with American society, through personal experience or that of friends and relatives, results in Mexican immigrants' greater awareness of the importance of knowledge of English for status attainment than that observed for Cuban immigrants. Their research findings indicate a strong motivation to develop English-language skills among Hispanics which stems largely from economic concerns. There is reason to believe that the English learned by Hispanics is actually closer to "school English" or more formal English than that acquired by many Blacks. If this is the case, the relatively higher attainment of the Hispanics on the verbal tests may be attributed to their having learned more formal and precise English skills.

It may also be the case that, if indeed a substantial portion of the Hispanic respondents are bilingual, their bilingualism facilitated their language test performance, particularly on the Word Knowledge test. Of the words included in this test, 69 percent have Spanish cognates (Foster, 1981). Foster, a professor of Spanish, notes that most "high class" English words have Latin roots that appear in the everyday language of Hispanics. The linguistically aware speaker of Spanish taking the Word Knowledge test would therefore have clues to the choice of the correct alternative available to few Whites and to even fewer Blacks.

The fact that a slightly higher percentage of the Hispanic respondents refused to take the test than was observed for the other groups (see Section 4.1) also helps explain these findings. It is likely that Hispanics with the poorest English skills are not included in this sample and that the average performance observed for this group is slightly increased, especially on the tests that depend heavily on English language knowledge.

To decide between alternative hypotheses about the Hispanics' language performance relative to Whites and Blacks or about the impeding or facilitating effects of bilingualism would require knowledge of the language development of individual subjects. The necessary information is not available in this study. The data suggest only that conventional ideas about English language as related to Hispanics in the United States may need revision.

## 5.2 Sociocultural Group by Region Effects

We pointed out in Chapter 4 that there are no statistically significant three-factor interactions of sociocultural group, economic status, and region. Neither is there any evidence of two-factor interactions of economic status and region. This means that the amount by which the test scores of the nonpoor groups exceed those of the poor is not noticeably different in the four regions of the country. Although, as we saw in Section 5.1, the effect of the economic status classification differs somewhat with years of schooling and sociocultural group, the general advantage of the nonpoor subjects over poor is about 20 points for any of the ASVAB tests.

The differences between sociocultural groups, however, interact substantially, and in interesting ways, with the classification by residence at age 14 (Region). Figure 5.2 shows the performance of the sociocultural groups in the Northeast, Southeast, Midwest, and West in terms of the standard scores of the national probability sample. (For a list of the region assignments of the states, see Table 2.4-6.) Although the patterns of these interactions show similarities when the tests are grouped by specialized knowledge, fluent production, vocabulary, and quantitative ability, these distinctions are less helpful in discussing the region effects and will not be used in this section.

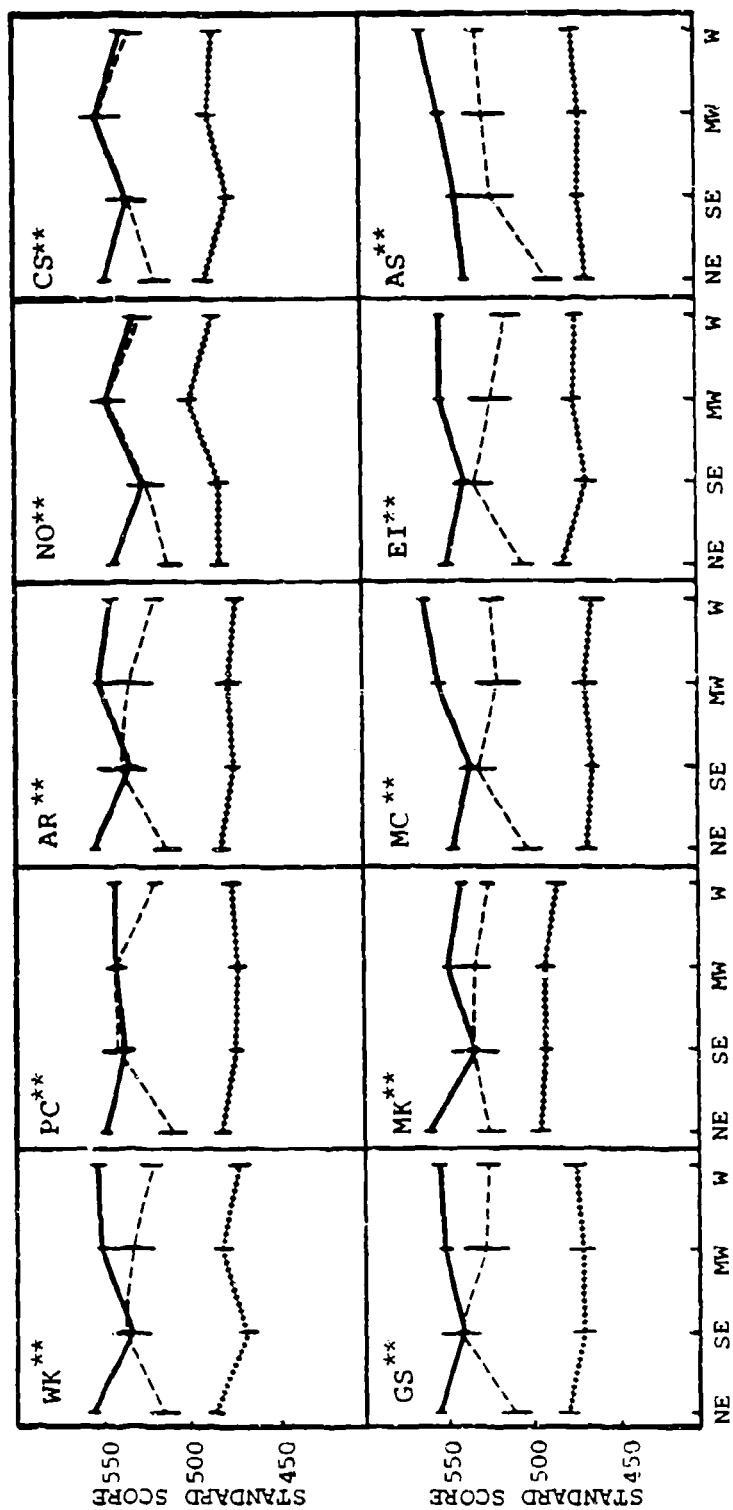


Figure 5.2 Interaction of sociocultural group and region of residence at age 14.

White——— Hispanic----- Black.....

\*\*F > 1.25F(.01)

Regional Effects on Hispanics' Achievement. By far the most conspicuous interaction of sociocultural group and region is that in the Hispanic subsample. The achievement of Hispanics in the Southeast region equals that of Whites in this region on seven of the ten subtests (i.e., Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, Numerical Operations, Coding Speed, Mathematics Knowledge, and General Science) and is only moderately below that of Whites on the remaining tests. Whites show somewhat higher average scores in the specialized knowledge areas of Electronics Information, Mechanical Comprehension, and Auto and Shop Information.

The performance of Northeast Hispanics is more similar to that of Blacks of that region on these tests. The substantially lower performance of these Hispanics in Numerical Operations and Coding Speed suggests that linguistic factors may not be the only variable in their lower attainment relative to Whites. Probably considerable numbers of Hispanics in each region are largely isolated from the majority culture. They would tend to persist in traditional cultural patterns as opposed to becoming bicultural. Ramirez (1977) has reviewed several studies which show that among Mexican-Americans, those who are bicultural (Mexican and Anglo cultures) are more successful at all levels of education from kindergarten to graduate school programs.

In contrast to the levels of achievement observed uniformly for the Hispanic groups of the Southeast and, for some tests, those of the Midwest, Hispanics in the West show considerably lower achievement than Whites in all the tests except Numerical Operations and Coding Speed. In the latter, the performance of Hispanics is nearly identical to that of Whites. Neither do Whites in this region greatly outperform Hispanics in Mathematics Knowledge, but the difference is significant.

The differences in the relative performance of Whites and Hispanics in the Northeast, Southeast, and West are probably due to the different social origins of Hispanics who reside in these regions. Given the predominance of Puerto Ricans in the Northeast, Cuban Hispanics in the Southeast, and Mexican Hispanics in the West, we are undoubtedly observing the effects of origins in different Hispanic cultures and socioeconomic classes. Many of the Cuban immigrants in the Southeast, especially those who came to the United States in the early 1960s after the revolution in Cuba, were highly skilled, well-educated business people and professionals who, with considerable assistance at the national, state, and local levels, were able to reestablish themselves economically (Pedraza-Bailey, 1982). Their children are now appearing in this sample. The work of Wilson and Portes (1980) indicates that the Cubans have been able to develop a strong economic base for their communities in the Southeast. In addition, Florida established a strong bilingual education program for the Cuban immigrants, using both state and federal funds, to facilitate immigrant children's and adults' education and adjustment to American life.

The Mexican-Americans in the West are not faring as well economically or socially as Hispanics of the Southeast. Hispanics in this group generally have come to the United States with few skills. In addition, the West experiences a remarkably high Mexican-American school dropout rate. The development of educational programs to facilitate Mexican-American children's school performance and adjustment to life in America has suffered from insufficient funds and fluctuating levels of larger community support. It is, then, not surprising that there are considerable differences in the relative performance of Hispanics and Whites observed for the western region compared to the southeastern region.

In the Northeast, which is heavily populated by Puerto Rican Hispanics, larger differences occur between the Hispanics and the Whites. These Hispanics score well below the level of Whites in this region on all the subtests. Generally their performance is closer to that of the Northeast Blacks, especially in Paragraph Comprehension and Auto and Shop Information.

We have not analyzed the performance of the Hispanic subgroups for this report because the resulting sample sizes would be too small for reliable analysis. However, Eitelberg and Doering (1982) have presented the average Armed Forces Qualification Test (AFQT) scores for self-identified Mexican-American, Puerto Rican, and Cuban respondents included in this sample. The AFQT is a composite score derived by combining the test scores of individual respondents on the Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning and Numerical Operations tests of the ASVAB. The AFQT score is one of the criteria used by the Military Services to determine applicants' enlistment eligibility. Though this composite score is different from the individual test scores we have analyzed to this point, it is suggestive of the relative scoring of these groups, particularly in the school-intensive skills. Eitelberg and Doering report that the Cubans in the sample achieve an average standard score on the AFQT of 494, Mexican-Americans achieve an average score of 448, and Puerto Ricans achieve an average score of 444. These scores are very similar to the pattern of scores we see for Hispanics in the various regions where the different Hispanic groups tend to occur.

Auto and Shop Information is the one test for which western Hispanics do about as well as or better than those in the other regions. Apparently, the automobile centered majority culture of the West, particularly California, extends to the minority Hispanic culture. Perhaps the "low rider" clubs can take some credit as well.

In the Midwest, Hispanics equal the performance of Whites in Paragraph Comprehension, Numerical Operations, and Coding Speed. On the Arithmetic Reasoning and Mathematics Knowledge tests they are about 10 points below Whites. On the other subtests they are 20 or more points below the performance of Whites. They are consistently about 50 points above Blacks on all tests. Why the Hispanics in the Midwest show their particular patterns of advantage and disadvantage in attainment relative to their White peers is not clear. It may be that in the Midwest, owing to the absence of any historical antagonism between Hispanics and Whites, the Hispanics are more fully integrated into schools and other community institutions than in the West. In the urban Midwest especially, social class has a greater influence than ethnicity. Alvarez (1976) observes that Mexican-Americans migrated to the Midwest and Northwest to escape the caste-like status they experienced in the Southwest. He points to the industrialization and religious, ethnic, and political heterogeneity of the Midwest as factors that attenuate discrimination against Mexican-Americans. The status of Hispanics in this region tends to be more one of lower class than of lower caste as in the Southwest. The stability of their tenure in the Midwest community may also account for this pattern, since unlike the situation in the Southwest, travel between the United States and Mexico is not as convenient. This Midwestern Hispanic community, composed of Cubans, Mexicans, Puerto Ricans and other Central and South Americans, is itself more heterogeneous. This greater diversity relative to the largely Mexican community of the West could also account for some of the regional difference in test performance.

Regional Effects on Blacks' Achievement. Blacks in the Northeast and Midwest regions show similar levels of performance on all of the tests and a higher level than Blacks in the Southeast and West. These data generally



confirm the earlier findings of Coleman and his colleagues (1966) that Blacks in the Northeast are achieving at levels higher than those of Blacks in other regions. The present data indicate, however, that Blacks of the Midwest are doing almost as well. This is especially true in basic skills.

Blacks in the Southeast generally show lower performance on the basic skills tests than those in the other regions, although the difference is remarkable only in Word Knowledge and Coding Speed. In the specialized knowledge areas, however, they equal the performance of Blacks in the other regions.

Unlike the regional effects observed for Hispanics, in no region of the country do the skills of Blacks approximate closely those of the White group. These data reveal clearly the pervasiveness of Black young people's lack of preparation for successful competition in the meritocracy. The factors that contribute to this failure, including ability grouping in schools, will be elaborated in later sections.

### 5.3 Sources of Sociocultural Group Differences

In the introduction to this chapter, we noted that all of the young people included in this sample were born or attained formal school age during an era of unprecedented efforts to provide equal educational opportunity to all American children. There was a national thrust to eliminate the educational handicaps that prevent minority status children from participating fully in the occupational structures of the society. We remarked also that if those good intentions had actually translated into effective action, we might have expected some positive effects to show in these Profile data. From this perspective, the findings of this survey are disappointing with regard to performance of Blacks, but are more encouraging for Hispanics and low-income Whites. The results raise important questions in view of the considerable

amount of time, energy, and resources which were focused on the problem of Black children's failure to benefit from schooling during the '60s and '70s. Why are the differences between Blacks' and Whites' average skill development observed by Coleman, et al. (1966) sixteen years ago still at about the same level? Some answers to this question will be suggested in the subsequent sections of this chapter. First we will focus on the general process of status attainment, consider the elements of the process on which Blacks and Whites may differ, and suggest how and why these differences contribute to variations in Blacks' and Whites' educational attainment and, ultimately, their occupational attainment.

The Process of Status Attainment. The place found by individuals in the socioeconomic structure of this country is strongly affected by educational attainment. People with more education obtain higher status jobs and, generally, earn more money than people with less education. Sociologists have developed a multivariate model of status attainment that has proved particularly effective in explaining the educational and occupational attainment process of White males and females (Duncan, et al., 1972; Jencks, et al., 1972; Sewell & Hauser, 1975; Treiman & Terrell, 1975; McClendon, 1976). Essentially, it is a five-stage model in which mother's and father's education, income, and occupational status, and the individual's ability (indexed by I.Q.), explain scholastic achievement. All background variables and scholastic achievement determine the influence of significant others (i.e., mother, father, teachers, best friends); all of the preceding variables affect the individual's level of ambition; and, finally, educational attainment and occupational status are determined by all previous variables in the model. Although the goal of much of the work on status attainment has been the development of a model valid for all society (Blau and Duncan, 1967;

Sewell and Hauser, 1972), comparisons of the educational attainment process of Blacks and Whites have shown that this model does not provide as good an explanation of the attainment process for Black Americans as it does for Whites (Porter, 1974; Portes and Wilson, 1976; Kerckhoff and Campbell, 1977).

Portes and Wilson (1976) observed differences in the educational attainment process of Blacks and Whites, particularly in the order of effects delineated by the model. Socioeconomic status variables, mental ability, and academic performance--variables appearing early in the model--were found to predict strongly overall White educational attainment. However, for Blacks, the strongest influence on educational attainment was observed for variables appearing later in the model--namely, self-esteem and educational aspirations. Similarly, Kerckhoff and Campbell (1977) report that family socioeconomic status is a much more important influence on the educational attainment process of Whites than of Blacks. These investigators conclude that the results for Whites often appear much more orderly and predictable (p. 25).

A similar model of status attainment has guided much of the thinking of psychologists and educators about factors that lower the school performance and educational attainment of minority status children. Cultural deprivation theory is an example. Bloom, Davis and Hess (1963) described culturally deprived students as:

. . . the students whose early experiences in the home, whose motivation for present school learning, and whose goals for the future are such as to handicap them in school work. This group may also include those who do not complete secondary education (p.4).

As this description suggests, these investigators believe that family background variables exert a strong influence on the entire educational attainment process. The specific focus of their work was the determination of

what early experiences resulted in these negative effects, and how these experiences mediated scholastic achievement.

But findings of Portes and Wilson (1976) and Kerckhoff and Campbell (1977) suggest that early experiences in the home may not be as critical to Black children's educational attainment as previously thought. Both of these reports suggest that the educational attainment of Black students may be determined more by what happens to them once they enter the school system than by what they bring to the system. However, they do indicate that I.Q. test scores significantly predict the educational attainment of both Blacks and Whites, though in each study the effects for I.Q. are stronger for Whites than for Blacks. Family background factors and associated early experiences may therefore be important to Black children's educational attainment to the extent that they mediate general verbal ability.

I.Q. Scores and the Educational Attainment of Black Children. Children from the lower socioeconomic levels of our society achieve lower average intelligence test scores than those from the middle-class (Jones, 1954; Lesser, Fifer and Clark, 1965; Golden, et al. 1971; Hall and Kaye, 1981). Investigators who have compared the intelligence test performances of different ethnic groups from both high and low social classes report significant effects within economic strata. Black children typically achieve lower average scores than their White peers when social class is controlled (Sitkei and Myers, 1969; Scarr, 1971; Nichols and Anderson, 1973; Trotman, 1977).

As Anastasi (1982) points out, I.Q. is the general designation of scores derived by tests designed to assess a number of different cognitive skills such as the ability to manipulate abstract symbols, verbal facility, numerical facility, and other such specific skills thought to be necessary to succeed in a particular culture. Since most of the current intelligence tests

were modeled after the original Binet Scale developed in France in the early 1900s for purposes of assessing the nature of children's learning difficulties in the school setting, tests today assess primarily verbal facility and to some extent, numerical facility, and reasoning skill (skills required by the school curriculum). The score obtained from this test, typically referred to as I.Q., is used to index both previous learning experience and to predict subsequent scholastic achievement. To the extent that it does the latter, the I.Q. score is thought to indicate "ability" or potential to benefit from future training of various sorts.

Because a positive correlation exists between intelligence test performance and school achievement (Cleary, et al., 1975), Black children's lower average I.Q. scores were early fixed upon as major factors contributing to lower scholastic achievement. Not considering the possible cultural origin of the mean I.Q. difference between Blacks and Whites, many researchers believed that if Black children's I.Q. scores could be raised, then their school achievement would also rise. Many evaluations of intervention programs designed to eliminate Black children's academic handicaps used performance on standardized intelligence tests to determine the effectiveness of the programs (Horowitz and Paden, 1973). Since Bloom (1964) had reported that I.Q. was relatively stable after the preschool years, it was assumed that if gains in Black children's I.Q. scores could be effected by intervention programs early in life, these gains would carry the child to higher educational attainment, and ultimately, higher occupational status.

While the success of these programs in boosting Black children's I.Q. scores varied depending upon their nature and structure, much to the investigators' consternation, the increments in Black children's scores were not maintained past the first few years of public school attendance. Jensen

(1969) theorized that the failure of the programs' efforts to produce lasting changes in Black children's I.Q.'s could be explained on genetic grounds. Jensen's theory prompted researchers to reconsider the relative contribution of genes and experience to I.Q., educational attainment, and vocational success.

Genes and Black Children's Intelligence. According to Jensen's (1969, 1973, 1980) theory, Black and White subpopulations differ sufficiently in the frequency of certain behaviorally predisposing genes to account for the typical one-standard-deviation disparity between their mean I.Q. scores. His reasoning is based primarily on research that shows substantial heritability of I.Q. scores within homogeneous populations. Heritability is an index expressing the proportion of the total variation of individual differences seen in a population for a given measurable trait that can be attributed to differences in inherited genetic constitution. Nonzero heritability of a trait implies positive correlation of the trait among relatives. Numerous empirical studies bearing on the heritability of I.Q. scores show that, as family relatedness increases, similarity in I.Q. scores increases. For example, Erlenmeyer-Kimling and Jarvik (1963) reviewed 52 studies which report steadily increasing correlation coefficients for I.Q. scores of individuals with varying degrees of genetic similarity (i.e., family relatedness). A more recent review based on many additional studies has been published by Bouchard and McGue (1981). The new data indicate somewhat lower coefficients of heritability for I.Q. scores than were previously accepted. The studies reviewed by Bouchard and McGue show a median correlation of .67 for monozygotic twins reared apart; while that of first degree relatives reared apart is about .23. Although these correlations are necessarily attenuated somewhat by unreliability of the I.Q. measures, they are also likely to be strengthened by

similarities of the environments in which separated family members find themselves. All things considered, the data reported in the literature suggest a heritability for I.Q. in the range of 50 to 60 percent. This would mean that 50 to 60 percent of the total variance in I.Q. seen in a population is due to heritable genotypic differences. However, the 40 to 50 percent of variation remaining leaves ample room for environmental influences to affect responses on these tests.

While geneticists agree that these findings indicate appreciable heritability of I.Q. within homogeneous cultural groups, they also agree that the typical methods and assumptions of genetic analysis regarding psychological test scores do not apply to comparisons of mean levels of performance between reproductively distinct populations (Thoday, 1969; Bock, 1974). Conventional models of quantitative genetics deal exclusively with variation within populations: intergroup comparisons of behavioral traits are problematic because, insofar as there is any environmental component in the trait variation, there is no basis for inferring the quantitative effect of environmental influences on the population mean. Jensen, in particular, violates this principle of genetic analysis when he extrapolates models of genetic variability and transmission derived from analysis of within-group variation of the trait (I.Q. score) to account for between-group differences in the population means. As a result, his conclusions regarding the genetic basis of differences between Blacks and Whites in I.Q. have no factual basis. Careful examination of this problem (Bodmer and Cavalli-Sforza, 1976) finds no grounds for Jensen's conclusion that racial and ethnic differences in intelligence are genetic in origin. In fact, there is good evidence to the contrary, as we shall see.

In order to justify between-group comparisons for genotypic I.Q. based on test scores, it is necessary to show that, within the two populations, environmental factors affect the development of intelligence in the same way and that complete identity in the distribution of the environmental variables exists between populations (Scarr, 1975). Even if the first point were conceded, the socioeconomic variability observed between Blacks and Whites, as well as differences in cultural patterns, does not allow the second point. Intelligence test performance is influenced not only by ability to apply memory and reasoning to a problem, but also familiarity with the cultural forms in which the problems are couched. Researchers now recognize that the intellectual and social isolation of Blacks in this country over more than a hundred years has either given rise to, or resulted in the maintenance of, substantial cultural differences between the White and Black populations. These differences are reflected in the way children are oriented to the social and object environment, modes of communication, and tolerance for varying levels of sensory stimulation (see Hale, 1980, for a review of this literature).

The proof of the correctness of this conclusion is in research showing that when Black children are reared in the majority culture, they achieve average test scores on I.Q. tests comparable to those of the White population. Scarr and Weinberg (1976) report that when Black children with at least one Black parent are adopted by middle-class White parents in the first year of life, they achieve an average I.Q. of 110. Of course, some might argue that since the Scarr and Weinberg sample contained a large portion of biracial children (i.e., children with one White parent and one Black parent), these children's White ancestry contributed to their higher performance. However, Moore (1980), in a similar study, reports that for Black children adopted by



middle-class White parents, there are no differences in average I.Q. score between Black children who are the offspring of two Black parents and those who had one White and one Black parent. The average Wechsler Intelligence Test for children (WISC) full scale score observed for the former group was 118, and for the latter group was 116.5.

Moore's (1980) study was specifically designed to determine the influence of the cultural environment on Black children's intelligence test performance, when socioeconomic status is controlled. She compared the WISC performance of two groups of adopted Black children: one group adopted by Black middle-class families (N=23), the second group adopted by White middle-class families (N=23). Both groups of children had been adopted from the same two agencies, and both groups of parents had met the agencies' stringent placement requirements. There were no significant differences between the Black children placed in the two kinds of adoptive homes in terms of biological mothers' health and prenatal care. All of the children were similarly healthy at the time of birth, and were comparable in the number and quality of preadoption placements, and age at placement (about 15 months for both groups).

In Moore's study, there were two differences between the Black children in the White homes and those in the Black homes. First, the average educational attainment of the biological mothers of the Black children placed in Black homes was significantly lower than that of the biological mothers of those placed in White homes. However, 15 of the biological mothers in the former group were still in school when their babies were born, and it was impossible to determine from agency records how much more schooling they actually obtained after the birth of their children. Second, biracial and "full Black" children were not normally distributed in the two kinds of

adoptive placements. There were more biracial children in the White homes (N=14), than in the Black homes (N=6). However, discussions with the social worker who made all of the placements for the two agencies indicated that this was not the result of the agencies' attempt to effect any type of environment-genotype correlation. Rather, this particular distribution of biracial and "full Black" children among the two adoptive placements was consequent to the Black parents' preference for a child who would resemble themselves in physical appearance.

Just as there was no significant difference in the WISC full scale I.Q. of biracial and "full Black" children in the White adoptive homes, there was no significant difference in the average performance of the biracial and "full Black" children in the Black adoptive homes. Biracial children averaged 105.7; the "full Black" children averaged 102.9.

The average full scale WISC I.Q. of all the Black children in the Black adoptive homes was 103.6; for the Black children in the White homes, 117.1. The 13.5 point score difference between these two groups of adopted Black children is of the magnitude observed between Black and White children in the general population. While the advantages of a middle-class rearing environment for Black children's intellectual performance are demonstrated in the scores of the Black children in middle-class Black homes, the advantages to Black children's test performance which come with rearing in a middle-class home of the majority culture are even greater. These data suggest the importance of the cultural component to Black children's I.Q. test performance, as well as the influence of a social class component.

Taken together, the findings of Scarr and Weinberg, and of Moore, indicate that when Black children are reared by parents in the majority culture, they obtain I.Q. test scores comparable to, if not greater than,

their White counterparts. Further, both biracial and "full Black" children benefit similarly from rearing in advantaged White homes.

To further explore the effects of racial genetic differences on individuals' intelligence test performance, Scarr, et al. (1977) analyzed the relationship between estimates of degree of White ancestry obtained for a large sample of Black and White children based on blood group markers (i.e., genetically determined antigens on red cells that have been found to be reliable markers of population differences) and their I.Q. test scores. Jensen (1973) hypothesized that there would be a correlation of approximately .50 between degree of White ancestry and I.Q. based on his estimates of the effects of genetic differences between Whites and Blacks for I.Q. test scores. However, Scarr, et al. report no significant correlations between their estimates of degree of White ancestry and children's scores on five different intelligence tests. The highest correlation found between these estimates and test performance was .13. This is very strong evidence that the mean difference in I.Q. scores of the Black and White populations does not arise primarily from genetic sources.

Genetic differences between groups are much more in characteristics that are adapted to climatic differences, such as skin color and hair form, than in behavioral or other physiological characteristics. Average intellectual attainments of populations that have been historically discrete are mostly expressions of their particular institutions and cultures, and are transmitted by imitation and instruction within the family, school, and community.

To summarize to this point, there is a growing body of evidence for a heritable component to I.Q. scores within populations, but no evidence to indicate that mean differences in I.Q. scores between exclusive populations

arise from this source. And, even within populations, where individual differences in I.Q. scores appear attributable in part to genotypic variance, environmental factors can strongly influence the performance of individuals on these tests.

Additional evidence of the influence that social/environmental factors can have on intelligence test performance within populations has been provided by Schiff, et al. (1982) from an adoption study in France. These investigators designed their research to address directly the question raised by Jensen in 1969, "How much can we boost scholastic achievement and I.Q. scores?". They identified a sample of French school children (N=32), abandoned at birth, whose biological parents were both unskilled workers, but who were adopted by upper middle-class families within four months of birth. The adopted children's I.Q. scores were determined by the administration of an individual intelligence test, the French version of the Wechsler Intelligence Scale for Children (WISC), and a group intelligence test, the ECNI, which, according to the investigators, is widely used in the French school system. The school records of the adopted children were also examined to determine their scholastic achievement. The effects of the change in the adopted children's social class membership on their I.Q. test scores and school achievement were estimated by comparing their performance to that of children of unskilled workers in the general French population which had been obtained in two previous large scale studies in France. An internal control group was also developed which consisted of the half-siblings of 20 of the 32 adopted children who had been reared by the working-class biological mother they shared.

Schiff, et al. report that the change in social class membership of the adopted French children increased their I.Q. scores by 14 points (about

one standard deviation) over what would be expected if they had been reared by their working class biological parents. The mean score observed for the adopted children on the group intelligence test was 106.8, for their half-siblings, 95.1, for children of unskilled workers in general in France, 95.1, and for biological children of upper middle-class families in general, 110. The mean WISC score of the adopted children was 110.6 as compared to a mean score of 94.2 found for their half-siblings on this individually administered I.Q. test, and an average WISC score of 100 observed in the French population generally.

The scholastic achievement of the adopted children was assessed primarily in terms of school failure rate. School failure was defined as either having to repeat grades or placement in classes for slow-learners. The school failure observed for the adopted children was 4/32, for their half-siblings 16/32, for children of unskilled workers in general 16.5/32, and for the biological children of upper middle-class parents in general, 4/32. Therefore, these investigators concluded that the change in social class membership of the adopted children reduced their probability of school failure by a factor of four.

The Schiff, et al. (1982) study is particularly well designed and executed. It provides very clear evidence of the significant influence of social variables on I.Q. test performance within populations, where the estimate of the heritability of I.Q. is in the 50-60 percent range. However, the findings of Schiff, et al. are not surprising. Geneticists recognize that environment can influence the average expression of traits even when they are highly heritable within populations. As a case in point, Bock (1974) cites the example of average stature in European, American, and Asian populations. He notes that in these three populations, measures of stature have a herita-

bility estimate almost identical to that of I.Q. scores. Yet, the mean stature of all three of these populations has increased nearly one standard deviation in two generations--an increase comparable to the observed difference in measures of intellectual ability between Whites and Blacks in the U.S. population. The increase in average stature of the Japanese since World War II is even more dramatic. This demonstrates that environmental influences (presumably better nutrition in the case of growth in stature) can substantially change the mean of a highly heritable quantitative trait. Noting the parallels between the heritability of stature and of general intellectual ability, Bock goes on to observe that:

. . . nothing now known about the determinants of general verbal ability should induce us to rule out a similar trend in the average ability of the U.S. Black minority as it moves out of years of segregated life and into the majority educational, linguistic and cultural environment of the United States (p. 595).

In short, the bulk of the evidence currently available does not support genetic explanations of the observed differences between Black and White children's I.Q. The explanation for the average differences in Black and White children's I.Q. and school achievement must be sought elsewhere.

The Schools and Black Children's Achievement. The theory which formed the basis of early school intervention programs is based on the assumption that experience in early infancy and childhood has a lasting effect on the individual (i.e., persists throughout the course of development). However, Clarke and Clarke (1976) and Brim and Kagan (1980) have compiled evidence from a variety of areas to demonstrate that this is not entirely true. These investigators conclude that the effects of early experience (both positive and negative) are not permanent, but that over the life-course, individuals change as a result of continual interaction between their capacities and their changing environments.

From this emerging view of a dynamic process of human development has come a new understanding of why Black children's I.Q. scores can diminish with public school attendance. The public schools have generally been unable to provide continued support for these children's intellectual development at the level provided by high quality preschool programs (Ramey & Haskins, 1981). Two decades ago educators recognized that the public schools were failing to educate Black children adequately, but it was assumed that the problems were with the children, not the schools. Now it is becoming more apparent that the major problem in Black children's scholastic achievement is that public schools are not attuned to their particular needs as learners. The majority of Black children may come to the school context with skills, attitudes, and achievement orientations that are different from those of middle-class White children (Cole and Bruner, 1971). These differences are interpreted as fixed deficits by many school personnel. The importance of cultural relativity for determining educational potential is generally recognized by theorists, but has not been sufficiently understood in the public school system. In the case of poor Black children, in particular, teachers do not believe that the children can master higher-level skills, and therefore do not provide the same level of exposure to new materials, encourage independent exploration, or require mastery at the level they do for children viewed as more able (i.e., middle-class White children) (Haskins, 1981).

I.Q. tests can be useful in providing an assessment of the scholastic skill development of Black children for purposes of diagnosis and remediation. However, I.Q. test scores are most often used to relegate Black children permanently to inferior positions in the school curriculum. When I.Q. scores are used to set limits on the type of education children receive with teachers providing instruction limited to the level of the child's I.Q.

and making only modest attempts to provide intellectual growth experiences, it is not surprising that I.Q. and subsequent educational achievement are so strongly correlated. There is evidence that, from kindergarten on, teachers make assessments, either formally or informally, of what each child can learn. They then teach to this level, which results all too frequently in Black children remaining behind their more advantaged peers as the schooling process continues (Rist, 1973).

Hall and Kaye (1980) present data that address this point. Citing results from a longitudinal study of 900 Black and White lower and middle-class boys, ages 6 to 9, who were administered tests of memory, intelligence, learning, and transfer over a four-year period, they conclude that at age 6 Black and lower-class children are behind their White middle-class peers in intellectual functioning and information processing. However, the Black and lower-class children, although they fail to "catch up," continue to progress at the same rate as middle-class children. These investigators interpret their findings to mean that at age 6 Black and poor children show a developmental lag in intellectual functioning behind their more advantaged age mates. They conclude, therefore, that formal school training should not be required of the former group until they are ready, so as to avoid poor achievement and an associated negative attitude toward school.

However, a plausible alternative interpretation of their findings is that, because at age 6 Black and poor children have lower scores than their more advantaged peers, teachers do not provide them with educational stimulation sufficient to support an acceleration of skill development to the point of "catching up." According to this view, teachers are, in effect, delaying training of these children. The negative affect which these children develop toward school may not be due to frustration in trying to master



school-intensive skills, so much as to frustration with the teachers' failure to provide appropriate instruction and exposure (Rist, 1973).

Scarr (1980), in commenting on the research presented by Hall and Kaye (1980), disagreed with their recommendation that formal school training should be delayed for Black and poor children until their underlying intellectual ability catches up to that of advantaged Whites. She points out that this results in their falling further behind in academic skill development. Since Black and poor children show developmental lag, Scarr recommends that they be given intensive training in basic verbal and numerical skills from kindergarten on (even if they do not get training in music, art, or recreational sports) to allow for the possibly longer time it will take them to master these skills. Certainly Black children must have these basic skills as prerequisites of the more complex abilities which will permit them to become competitive in the world of work. Scarr's recommendation is a reasonable one in light of the current findings.

Fernandez, et al. (1975) report from their study of high school students that Chicanos and Blacks have a higher self concept than expected self-concepts for students low in achievement. These investigators attribute this finding to the fact that teachers set lower standards for students who enter high school with below average test scores, and reward them with grades and praise for minimal achievement. As a result, the students assume that they are making appropriate progress in school because they do not get feedback to the contrary. To illustrate the problem of school standards for these young people's achievement, these investigators report that even though Black and Chicano students had low skill levels at high school entrance, teachers did not provide them with assignments sufficient to challenge them. The students with the lowest achievement scores were also more likely than other

students to say that the assignments were "just right" for them. In contrast, Asians and Whites, who tended to have higher achievement scores, more often reported their assignments were too difficult. Further, among the students with low achievement scores, those who indicated that they spent fifteen minutes or less on homework per week were as likely to report that they usually completed their homework as not. These investigators conclude that high standards were not being communicated in these schools to low achieving students.

Mackler (1977) has reviewed a number of studies of minority status children, particularly Black children, who are successful in school and/or score high on I.Q. tests. These studies all come to similar conclusions. In one of these studies, McCable, et al., identified school children living in East Harlem (New York City) who had full scale WISC scores of 112 or over, a verbal scale score of 110 or over, and a performance scale score of 115 or over. From nine elementary schools in this area achievement test scores were obtained for sixty-seven children who met these standards from the 2nd through the 4th grade (thirty-seven Black children; twenty-three Puerto Rican children; and seven White children). Despite these children's high I.Q. test performance their achievement was well below what would be predicted. An example of their underachievement is seen in their performance on the Stanford Achievement Test. Of the sixty-seven children, twenty-one were one or more years above level, twenty-one were less than one year above level, two were on grade level, and twenty-three were below grade level. According to Mackler, the clinical summary of this group states that

"... though well-endowed, well-behaved, and receptive to intellectual stimulation, the children's performance was below their capabilities. Both teachers and parents were too easily satisfied by the children's modest attainment . . . . Parents' interaction with the schools was limited . . . ."  
(1977, p. 11)

According to this point of view the diminished skill development in Black children is the result of cumulative educational deprivation. Ogbu (1978) concludes that many schools fail to prepare Black children for desirable adult roles to the same degree as White children because the schools passively accept the cultural stereotypes that pervade American society. The education by sociocultural group interactions in the Profile data support the conclusion that Blacks are not receiving the same benefits of education as Whites, and the literature implicates the schools and other institutions in the community in this condition of undereducation.

This assessment reflects the situation current at the end of the 1970s, but it does not take into account trends in achievement of Whites and Blacks during this period. Evidence of change in relative levels of attainment in a number of school subject matters over this period has been investigated by Burton and Jones (1982). Using test scores of national probability samples of children age 9 and 13, as collected by the National Assessment of Educational Progress, they examined differences in average percent correct responses of Whites and Blacks to test items in writing, science, mathematics, social studies, and reading. Between 1970 and 1980, differences between Whites and Blacks declined significantly in all of these subject matters, except for an isolated instance in science in 1970 and 1972. The decreasing differences were highly consistent, occurring in both of the five-year intervals in which the tests were administered, and equally in the 9- and 13-year-old age groups.

Although these reductions in the gap between the scholastic performance of Whites and Blacks are modest in size relative to the size of the difference, the fact that they seem to be continuing is reason for optimism concerning the eventual equalizing of educational opportunity and

attainment. There is nothing in these data to support those who allege that efforts during the 1970s to enhance access to quality education have failed; on the contrary, there is every indication that Blacks, and presumably also other sociocultural groups, will improve their relative standing in school attainment as the progress of education in the United States is assessed during the 1980s.

#### 5.4 Summary and Conclusions

The Profile study presents an accurate and detailed picture of average vocational test performance in 1980 of the three main sociocultural groups in the United States, namely, Whites, Blacks, and Hispanics. For the ten tests of the Armed Services Vocational Aptitude Battery it reveals how performance of these groups varies with educational level, economic status, and region of the country. Relative to the total range of test scores, variation from these sources is large, and its very existence prods us to attempt an explanation of the differences seen in graphical summaries in Tables 5.1 and 5.2. As we have discussed in Section 5.3, theories as to the origin of these differences already exist in the scholarly literature. In the present section we review our conclusions that three of these theories have little or no support and suggest a fourth theory that we believe to be closer to the truth.

Genetic Endowment Theories. In Section 5.3 we documented the fact that books and articles can still be found that invoke theories of more and less favorable genetic endowments as the explanation of behavioral differences between exclusive populations. But we also cited four independent studies the results of which clearly exclude this possibility for measures of general ability.

One of these studies demonstrated that estimates of the proportions of the genetic make-up arising from the source populations are uncorrelated to

general ability. Such correlation is a necessary consequence of a genetic theory of population differences in the trait, and failure to find it effectively rules out the theory.

But perhaps more telling are the cross-fostering studies, also referred to in Section 5.3, in which children from disadvantaged social groups are adopted at an early age into middle-class White families and raised totally within the White sociocultural context. Such children are found to have the same average scores on general intelligence (I.Q.) tests as children in the White population. The same has been found to happen when lower-class White children are adopted and raised in middle-class homes: their average scores are at the level of the middle-class population and not that of the working-class population into which the children were born. These studies show that the demonstrable tendency of performance on I.Q. tests to be partly heritable within homogeneous populations fails to explain the differences in average performance between exclusive populations. Such differences appear to be concentrated in that part of I.Q. variation that is controlled by environmental factors.

Formal Models of Status Attainment. Although formal quantitative models have been constructed to account for variation in educational and occupational status, their interpretation is questionable when they use an I.Q. measure as an explanatory variable. Even when these models successfully describe relationships between status and other variables, without explaining the source of the I.Q. test performance differences, they do not identify the ultimate causes of the sociocultural group differences. We must therefore seek elsewhere an explanation for the differences we see in the Profile data.

Linguistic Theories. Another theory cogently argued in the literature is what may be called "the linguistic hypothesis." Some researchers suggest

that because most test performance and much of school attainment depends upon proficiency in middle-class English, differences observed between sociocultural groups reflect disparities in the quality of language (Bernstein, 1961).

There can be no denying that the major sociocultural groups have their distinct variants of English, but it is not clearly established that these differences are important in scholastic attainment or test performance.

The Profile data provide some insight into this question. Insofar as differences between the White and Hispanic groups are concerned, the Profile data suggest that a linguistic factor is involved to some extent. The evidence is the almost complete lack of differences between these groups on the Numerical Operations and Coding Speed tests. Proficiency on these tests is largely independent of language; the cognitive operations required by the tests can be performed in whatever language is convenient for the examinee without interfering with correct response. Only when we examine the tests that require English language proficiency, such as the Word Knowledge and Comprehension tests, do we see a notable difference in the performance of Hispanics and Whites. But the size of the difference is surprisingly small--less than the difference between Whites and Blacks on these tests. The indication is that the linguistic factor is not a salient determinant of the ability of Hispanics to respond to tests in this battery. Larger differences are seen for such tests as Electronics Information, Mechanical Comprehension, and Auto and Shop Information that require more specialized technical knowledge.

The Profile data therefore suggest that a difference in educational emphasis is more important than the language factor in determining the performance of Hispanics on these vocational tests. The majority of Hispanics have not yet moved into those types of occupations for which specialized technical

knowledge is a prerequisite, and their preparation is more limited in these areas. Further evidence for this conclusion is the fact that in the Southeast region, where Hispanics participate more broadly in the economy and enjoy a greater variety of occupations, differences between Whites and Hispanics are almost nonexistent on all of the tests. The Profile data lend some support to the linguistic hypothesis vis-a-vis the White and Hispanic populations, but implicate educational effects equally or more so.

In respect to average performance of the White and Black sociocultural groups, however, there is little if any evidence of a language effect. Differences between the group means for the non-language Numerical Operations and Coding Speed tests are almost as large as those for the tests that place heavy demands on language proficiency. As we have seen, there is an extraordinary uniformity in the gap between average performance of Whites and Blacks on the ASVAB tests. It appears at all education levels, for both poor and nonpoor economic classes and in all regions of the country. It is not purely a regional effect reflecting differences in urban and rural backgrounds. In fact, performance of Blacks varies less across the four regions of the country than does that of Whites and Hispanics. It is not limited to lower education groups but also appears among the participants of the Profile study with 13 or more years of school. With minor exceptions, the difference in means between Whites and Blacks is the same for those classified "poor" as it is for those classified "nonpoor."

Early Deprivation Theories. These latter observations bear on yet another theory of group differences to be found in the literature--the theory of cultural and educational deprivation, especially deprivation in the first six or eight years of life. This theory holds that the scholastic performance of disadvantaged children is impaired in the earliest years of schooling and

becomes even more noticeable later because of the cumulative effect of generally slower rates of learning. The Headstart program of early childhood educational enrichment is based on this theory.

We have already noted that subsequent researchers have questioned the deprivation hypothesis in its original form and have suggested that levels of educational achievement are affected throughout a person's formative years. Gains in early childhood will not necessarily be retained later; nor is it impossible that educational deficits will be made up.

Moreover, the Profile data do not seem to be in accord with an early deprivation theory. First, if the disparity between Whites and Blacks is due to educational and economic impoverishment, then it should be smaller for the nonpoor group than for the poor. In fact, the difference tends to be larger for the nonpoor. If there is a deprivation effect, it must therefore be affecting poor Whites more than Blacks. Second, we should see reduced differences between Whites and Blacks in the nonpoor 13+ (some college) groups. We would have to expect that Blacks in this more educationally select group are relatively less deprived than, for example, those who have no more than 8 years of education. But again, the differences tend to be largest in the college group and smallest in the 8-year group, so that if there is a deprivation effect, it is acting more on Whites than on Blacks. The pattern of average scores for the sociocultural groups at these four educational levels does not suggest a deprivation effect, unless the term is used so broadly as to be synonymous with uniformly lower levels of scholastic attainment and test performance.

The Community Norm Theory. We believe that the differences seen in Figures 5.1 and 5.2 between the mean scores for Blacks and Whites have a simpler explanation, hinted at in the studies referred to in the final para-



graphs of Section 5.3. These studies show that children from the urban Black communities who display relatively high verbal I.Q. scores (110 or better) do not have levels of school attainment comparable to children of this I.Q. range from White middle-class communities. The difference in attainment, which is too large to be attributed to regression to the mean for a highly reliable test such as the WISC, suggests rather a lack of standards and challenge that stimulates a higher level of attainment. Schools are charged with teaching those kinds of knowledge and skills that the child can not learn spontaneously from his surroundings outside the school. They do this by setting standards of performance, especially performance on tests, and challenging the pupils to reach those standards with the aid of the instruction and learning resources that the school offers. To the pupil it appears that the standards are set by the classroom teacher, and in part they are, but no one teacher can make demands on his or her pupils that too greatly exceed those of other teachers in the school. This is because the school is a kind of social system characterized by norms and expectations for student attainment and behavior that largely determine the instructional activities and performance standards of teachers. These points have already been made by Brookover et al. (1979). All members of the school social system participate in the development and maintenance of the school's normative climate and find ways to constrain performance demands from individual teachers that are perceived as inappropriate according to existing norms.

Similarly, no one school can set standards of attainment too far above those of the community in which the school functions. Unless the school is highly selective in admitting students, the community norm will prevail and the higher standards will not be met. This situation seldom arises, however, because teachers and principals of local schools seldom strive to set an

exceptional standard. Moreover, economic exigencies bar schools from attempting to rise much above the community norm. Meeting higher standards requires extra expenditures for better qualified teachers and enhanced instructional materials and facilities. In the face of limited budgets, it is more often the lowest tolerable standard that is met rather than an ideal and seemingly utopian goal. So the compromise between cost and quality is made more or less on the same terms from year to year, the status quo maintained, and the historically prevailing standards perpetuated.

The effect of school standards will be most clearly seen in knowledge and skills that are learned primarily through formal schooling. A good example in the ASVAB test is Arithmetic Reasoning. Most of us know how to do quantitative problem solving only because we have been taught specific strategies and algorithms for their solution. We do not learn to compute volumes and solve ratios and proportion in casual activities outside of school. So it is this sort of test that shows especially large sociocultural group effects and increasing differentiation of the groups at higher levels of education. The same is true of the other specialized technical knowledge tests.

This is not to say that learning outside the school is not also subject to social norms. Each family implicitly or explicitly sets standards for many of the aspects of attainment and behavior of its children. As we shall see in Chapter 6, the mother necessarily plays a prominent role in this process. But because children influence and are influenced by their peers, no one family can enforce a standard too high above that of the community. Indeed, many families may have no clear conception of a standard other than that which emerges by an informal consensus out of the typical behavior patterns of friends and relatives. (See Blau, 1981, for a discussion of the

significance of the social milieu of families as distinguished from SES, for children's scholastic achievement).

If we view individual behavior as largely an adaptation to the demands of a person's immediate social environment, we realize that children, whose horizons are limited, respond primarily to the norms of the community in which they play their part. These norms determine most of their manner of speech, conduct, interests, aspirations, and motivations. The average effect of such determinants of behavior can be greatly different in communities that, for historical reasons, have been separated from the majority culture by barriers of communication and cultural interchange for many generations. In the United States the major sociocultural groups, and perhaps to a lesser extent the smaller ethnic communities, have existed under conditions that have encouraged the development of different norms.

In the case of a test such as Coding Speed, for example, it must be this broader social process, involving all of the community and not just the schools, that explains the difference in performance between Blacks and Whites. The Coding Speed test, which has no formal educational content beyond the ability to recognize words and numbers, is mainly an exercise in organizing and performing a simple task as rapidly as possible. The examinee's level of motivation is critical, as well as his or her typical pace of doing any sort of paper and pencil task. That Blacks and Whites with 13 or more years of education should differ considerably in their average performance on this test can only be understood as differences in norms in responding to this arbitrary kind of task. Blacks and Whites must interpret differently the test instructions to mark the items "as fast as you can."

Although we have cited studies that cast doubt on a rigidly interpreted theory of early deprivation as the source of the sociocultural group

differences, we would also take the somewhat similar position that a community standard, once internalized by a particular person, is not easy to change. Thus, the most dramatic effect of change of standards is seen in the cross-fostering studies discussed above, where children were moved entirely out of one community and into another early in life. To effect such a change in an older child would require more than a short-term intervention. In young people and adults such a change would be even more difficult. This would explain why the average performance of Blacks in the service training schools, discussed in Chapter 3, is in accord with their lower average scores on the ASVAB tests. The training school is a kind of intervention, but it is too limited to make a profound change in habitual modes of responding to training and instruction.

The community norm theory also helps us to understand the regional differences in the ASVAB test performance of the Hispanics. The different groups of Hispanics in the United States, having migrated from different parts of Latin America, and having brought with them norms typical not only of different countries but also of different economic strata, maintain typical levels of average response because of the inertia within communities that makes institutions function in much the same way from one year to another. The differences reinforce themselves: higher levels of attainment bring the resources that make such levels easier to maintain; lower levels deprive communities of the skills and leadership needed to create a higher standard.

If the community norm interpretation of sociocultural group differences is correct, it would suggest that schools have a special role to play in raising community standards. The school is the main interface between the larger society and the local community. It is the avenue by which young people can be introduced to a broader view and higher aspiration. It can

challenge the community to a more stringent standard--if not in the present generation then in the next, as the youth begin to take their place in the work and leadership of the community. The response to such initiative may be quite slow, measured in decades and generations, but the ultimate consequences in terms of the economic health of the nation and the well-being of all of its people are so great that that effort has to be made. And the effort must especially be made in those communities where for too long very modest levels of attainment have been accepted as the norm, with little hope or belief that they can be changed.

Just such an effort has been underway in the Atlanta, Georgia, public schools since about 1972. Alonzo Crim, the superintendent of Atlanta public schools, outlined a plan in 1981 which he projected would help pupils in this public school system to achieve at the national norm in basic numeracy and literacy skills by 1985. Crim described this as a "bold goal" in view of the fact that the great majority of the students served by the Atlanta public schools are Black and poor, the school system is perceived by the larger community as low-achieving, and the pupils in the system have scored well below the national norms on standardized tests of achievement since the testing program was introduced into the system. Central to the plan that Crim advanced is the development of what he terms a "community of believers" in the ability of children to meet this goal (i.e., a change in the community norm of student achievement), which he views as a critical step in building pupils' commitment to learning and attainment. With the support of students, parents, educators, business persons, members of the clergy, the media and the citizens-at-large, Crim reported steady progress toward the 1985 attainment goal between 1980 and 1982 (1981, 1983) and achieved this goal in 1983--two years early (personal communication). The remarkable gains seen in the

academic achievement of Atlanta public school children in a ten-year period attest to the significant influence of the community norm on student achievement and the role that educators can play in helping to upgrade the norm of achievement that characterizes the communities which their schools serve.

That these and similar programs of educational enrichment aimed primarily at disadvantaged children are beginning to have a positive effect is documented at the end of Section 5.3. Analysis of surveys of school achievement carried out during the 1970s by the National Assessment of Educational Progress reveals that differences between Whites and Blacks age 9 and 13 are steadily decreasing. If these trends can be maintained by consistently applied programs to raise community standards and enhance the quantity and quality of education, the sociocultural group differences seen in the Profile study of 1980 should decline substantially over the next 10 or 20 years.

## CHAPTER 6

### THE EFFECTS OF MOTHER'S EDUCATION AND RESPONDENT'S AGE AND EDUCATION

We have seen in Chapter 5 the effects of cultural, economic, regional and educational variations in young people's backgrounds on their ASVAB test performance. These phenomena are, however, merely the end-effect of processes initiated by myriad factors in the child's immediate environment. In this chapter we attempt to make the case that mother's education, which is among the best predictors of the cognitive characteristics of her offspring at maturity, is one of the most important of these factors.

In addition, we also attempt here to disentangle the effects of the respondent's age and education on the development of the skills assessed by ASVAB. To a great extent, increasing years of education are confounded with increasing age because legal requirements for young people's school enrollment and attendance tend to create educational cohorts from birth cohorts. But, because of differences in grade promotions and individual decisions to leave school prior to secondary school graduation, educational cohorts and birth cohorts in the Profile are not identical. We therefore have some information in the data about age and education that enables us to estimate them separately. Our analysis of age effects is based primarily on summaries of the Profile study showing the test performance of persons of different ages who have completed the same number of years of schooling.

#### 6.1 Mother's Education

Most of the previous research on the influence of family-background variables dealt with intellectual ability and used I.Q. tests, which measure mainly general verbal skills. The Profile data go considerably beyond this, however, and also measure more specific attainments. We are therefore able to

examine the effects of one of the most important of the family-background variables, mother's education, on a wide range of skills.

#### 6.1.1 Results of the Profile Study

Although we see highly significant main effects of mother's education for each of the ASVAB tests, we find, somewhat to our surprise, only weak interactions with the other background variables. Earlier research literature would suggest appreciable interaction effects between mother's education, sex, and sociocultural group (Loasa, 1983; Featherman, 1980). But we have found only minor effects, involving most prominently Auto and Shop Information and to a lesser extent the other specialized knowledge tests. For most of the tests, mother's education has only the highly regular main effects plotted in Figure 6.1-1. Generally, as mother's education increases, the mean performance of the respondents increases, but the strength of this trend varies from one test to another.

Fluent Production. In the two tests where females have some advantage over males--Numerical Operations and Coding Speed (see Chapter 7)--mother's education is seen to have a relatively small effect on Coding Speed scores. For Coding Speed, the mean difference between respondents whose mothers have 0 to 8 years of education and those whose mothers have earned at least a college degree is about 60 points. But having a mother with 16 or more years of education gives only a slight advantage in Coding Speed score over having a mother with a high school diploma. Apparently, there is a limit to the influence that a more favorable family environment can have on this simple performance test.

A somewhat different pattern is seen for Numerical Operations. The performance advantage of respondents whose mothers have college degrees or more education, relative to those whose mothers have no more than an 8th grade



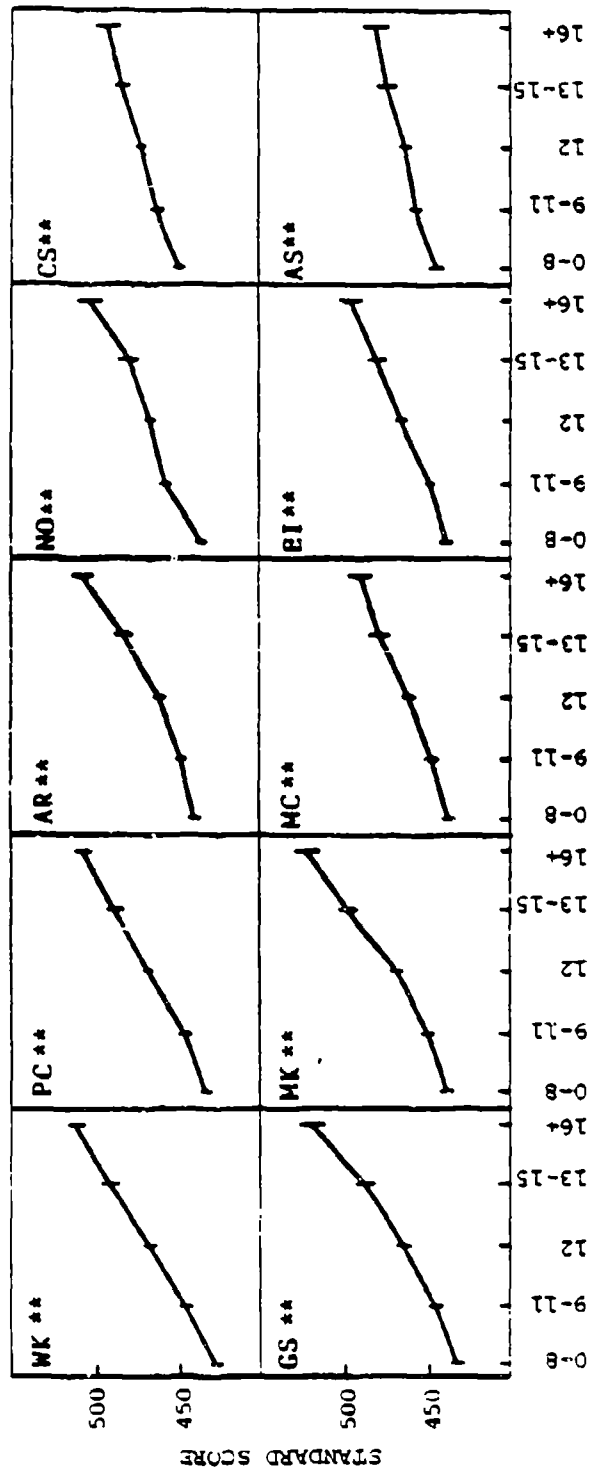


Figure 6.1-1 Effects of mother's education (mother's highest grade completed)

\*\*  $P > 1.25 P (.01)$

education, is modestly larger at 75 points. But unlike the effects for Coding Speed, having a mother with 16 or more years of education greatly enhances respondents' performance on Numerical Operations. A possible interpretation is that the Numerical Operations task, which all children are expected to master, is more sensitive to mother's press for children's scholastic attainment than is the more recondite coding speed task.

Specialized Knowledge and Quantitative Attainment. The relationship of respondent's performance to mother's educational level is similar and strong for all the more academic topics--Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, General Science and Mathematics Knowledge. The effect is greatest for Mathematics Knowledge, where respondents whose mothers have college degrees outperform those whose mothers are in the lowest educational level by a full standard deviation. As mother's education increases from some high school to the highest educational level, there is a steady and substantial increase in respondents' performance on this test.

At the other extreme, the young people's performance on Mechanical Comprehension, Electronic Information and Auto and Shop Information is not affected nearly as much by mother's education, although the trend is still positive. The least affected is Auto and Shop Information: changes in mother's education are accompanied by only small increases in Auto and Shop Information. Since this subject matter is more likely to be learned through outside experience than in school, the mother's education may not be as important. In other words, mother's education appears to be most important for attainment in school-intensive skills. The effects of mother's education on Mechanical Comprehension and Electronics Information are very similar. For both these subtests, the difference in the performance of respondents whose mothers have a college degree and those whose mothers have an 8th grade education or less is about 75 points.

There is a steady increase in average performance on all the tests with mother's education, but the amount of gain to the respondent is less for the specialized knowledge tests than for the more academically-oriented tests. We may be seeing here the effect of the learning ability inherited from the mother and achievement values acquired from her, but not necessarily the direct contribution of knowledge and instruction that the mother could offer in less technical areas where she is likely to have some attainment if she is sufficiently educated.

Note, however, that for the specialized knowledge tests, the mother's education main effect in Figure 6.1-1 is not a complete description of the variation. There is in addition some degree of interaction with sex and with sociocultural group. The interaction with sex, most apparent in the Electronics Information test, is plotted in Figure 6.1-2. There is a small but regular tendency for sons' scores to increase more than daughters' as mother's education increases. An obvious, and quite satisfactory, explanation would be the effect that sex-role expectations have on how young people apply the advantage an educated mother brings to the home in the development of these specific skills. Male offspring, for example, with their greater interest in technical material, may be more likely to apply the advantage an educated mother brings to the home to the study of electronics than would female offspring. The size of this interaction seems, however, too small to be of any practical importance.

The interaction of mother's education with sociocultural group, shown for General Science and Auto and Shop Information in Figure 6.1-3, is more interesting. It appears that performance on these tests is more strongly dependent on mother's education in the Hispanic group than in the other two groups. The group effect is concentrated mainly in the range from only-high-school education to college-graduate.

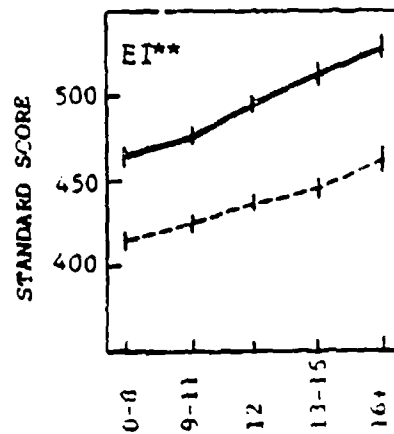


Figure 6.1-2 Interaction of mother's education (mother's highest grade completed) and sex. Male — Female - - - -  
 $**F > 1.25 F(.01)$

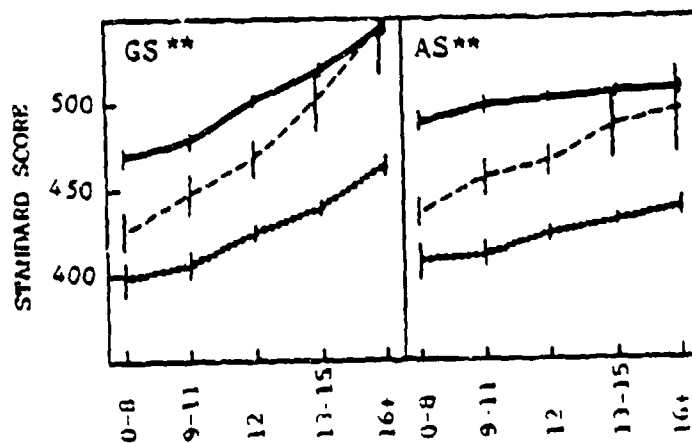


Figure 6.1-3 Interaction of sociocultural group and mother's education (mother's highest grade completed). White — Hispanic - - - - Black .....  
 $**F > 1.25 F(.01)$

Among offspring of mothers with only-high-school education, the Hispanics are about 30 points below Whites, but among those whose mothers have college degrees, Hispanics and Whites perform at essentially the same level; Hispanics are slightly higher than Whites in General Science and Whites slightly higher than Hispanics in Auto and Shop Information. That this effect is seen mainly in the specialized knowledge tests suggests a press for achievement in homes with better educated mothers that is turned more towards scientific and technical pursuits in Hispanic homes than in White or Black homes. This orientation is suggested by Brown et al. (1980), who report that a higher proportion of Hispanic youth who pursue postsecondary educational programs are enrolled in noncollegiate schools than in four-year colleges. In addition, a higher proportion of Hispanics in noncollegiate programs are enrolled in trade and industrial technologies than is true for Whites enrolled in similar institutions. Astin (1982) also reports that a higher proportion of Hispanic students in four-year colleges identified themselves as engineering majors in their freshman year of college in 1971 than was true for Blacks or Whites, though the proportion who actually obtained baccalaureate degrees in engineering was nearly identical to that of White students.

As we have pointed out in Chapter 5, much the opposite is true of Blacks. In the absence of abundant role models in the scientific fields, Blacks from homes with well-educated mothers look more to the nontechnical fields and prepare themselves accordingly. As a result, the trend for increased scores on the specialized knowledge tests with increasing mother's education is weaker among Blacks than among Hispanics and Whites. Compared to the main effect of mother's education, however, the difference in response to the specialized information tests among the sociocultural groups is relatively small.

Summary of Mother's Education Effects. We attach special significance to the fact that mother's education interacts only weakly with the other background factors. We take this as an indication that specific familial effects, genetic and cultural, act in a summative manner with those general effects of school and community that impinge on all subjects. Of course, the mother's educational level is itself a response to both familial and community influences, but the contributions are linear, not multiplicative, and no interaction of the two types of influence appears. That such effects are simply additive justifies in a certain way the separation of the disciplines of developmental psychology, which concentrates on the intra-familial influences, from social psychology and sociology, concentrating on the extra-familial influences on behavior and attainment. Taken together, the present data indicate that the familial effect indexed by mother's education is greatest on the offspring's academic skills represented in the ASVAB by Word Knowledge, Mathematics Knowledge, General Science, Arithmetic Reasoning, and Paragraph Comprehension. The mother's educational level has a smaller effect on the respondents' achievement on the fluent production items (Coding Speed and Numerical Operations) and on technical skills (Mechanical Comprehension and Electronics Information). The smallest effect of all is seen on Auto and Shop Information. In general, those young people who have mothers with a college degree have developed higher skills in all areas tested than those whose mothers have less education. Young people whose mothers have 8 years or less education show the lowest level of skill development in all areas. While having a mother with some high school increases the offspring's performance, the largest gains are seen among respondents whose mothers have achieved a high school diploma and a college diploma.

The general pattern of the effects of mother's education is consistent with the existing literature on the relationship between mother's socioeconomic status (which uses educational attainment as an index of social status) and her children's general ability (measured, for example, by I.Q. tests). In the present data, the advantage of having a highly educated mother is pervasive for offspring's performance across all tests. The fact that mother's education has a strong effect on offsprings' Word Knowledge is also consistent with the empirical and theoretical literature. Females' general superiority in language (Maccoby and Jacklin, 1974) is probably amplified at the upper educational levels, resulting in more verbally complex environments for individuals with highly educated mothers. These data generally support Williams' (1976) observations that highly educated parents tend to provide more complex learning environments for their children and encourage academic excellence.

#### 6.1.2 The Process of Mother-Child Interaction

Although there is abundant evidence that the socioeconomic status of the family, as indexed by mother's and father's education, occupation and income, is strongly related to offspring's educational and occupational attainment (Jencks, 1972, 1979), the specific factors that mediate the action of this complex of demographic variables remain obscure. One finds in the developmental literature a body of research based on the theory that the effective agent of these macroenvironmental factors is the early mother-child interaction. Much of the effort of developmental psychologists has been to identify in families of different social strata the processes of mother-child interaction that contribute to social class differences in early cognitive development, achievement orientations, values and learning styles (Witkin, et al., 1962; Hess and Shipman, 1965; Crandall, et al., 1960; Wolf, 1964; Chan

and Rueda, 1979; Farran, et al., 1980). In large part, their work has been stimulated by the observations that individual and group differences in these aspects of behavior appear early in life, prior to systematic experiences with cultural institutions beyond the family. This suggests that the family is the dominant conveyor of cultural conventions and that mothers, who in all societies are invested with early caregiving, are the primary mediators of children's early development of skills, attitudes and beliefs. If recent trends continue, the mother's influence in this role can only increase in the United States as the number of children reared in mother-only families grows with the increasing rate of divorce and the rise in the number of children born to never-married women (Glick, 1979). Because of all these influences, cultural variations in mothers' attitudes, beliefs, expectations and aspirations regarding their children may be expected to leave their mark on vocational test scores in the form of group differences in performance.

Conversely, the socioeconomic status of the family influences the pattern of maternal behaviors that in turn affect children's intellectual and emotional development as well as school success. Family socioeconomic status has been found to be related to the amount, variety and complexity of the stimulation that mothers provide (Dave, 1963; Tulkin and Kagan, 1972; Tulkin and Covitz, 1975; Jordan, 1978). It is also related to the style of interactions in the home, especially language behavior and maternal control strategies (Bernstein, 1961; Streissguth and Bee, 1972; Blau, 1981), and to maternal teaching strategies (Hess and Shipman, 1965; Hess, et al., 1969; Loasa, 1982). All of these behaviors appear relevant to children's cognitive and affective development, both by shaping intellectual processes and affective modes in children during the early years and by inducing a motivation to learn (Slaughter, 1983).



Many sociologists aggregate the mother's educational attainment with other indices of family socioeconomic status (i.e., parents' occupational and income attainment) to develop a global index of social class. However, some investigators believe that parental education in general, and mother's education in particular, is the most salient of these sociodemographic variables in determining learning strategies and values related to later school success (Hess, 1970; Jordan, 1978; McClendon, 1978; Willerman and Stafford, 1972; Loasa, 1982). Loasa (1982) theorizes that the high frequency of scholastic failure observed among young people from minority status groups may be explained largely by the fact that their parents, particularly their mothers, have generally completed fewer years of schooling. Loasa's conclusions are drawn from a series of comparative studies of intact Anglo- and Mexican-American families and their preschool children specifically designed to disentangle the effects of culture, parental education and parental occupation on children's scholastic attainment. His data indicate that the amount of schooling completed by the mother is a major determinant of both the press for verbal skill development in the preschool years, and of the mother-child interaction processes in the home that provide children with the opportunity to master "the forms and dynamics of teaching and learning processes" that have adaptive value in the classroom (Loasa, 1982, p. 800). Neither mother's nor father's occupational status was found to have a significant relationship to any of the parental behaviors examined in the Mexican-American subsamples; in the Anglo subsample, father's occupation was related only to maternal aspirations for the child's educational attainment.

Value orientations and community norms appear to underlie many of the differences in socialization observed between families that differ in amount of parental education (Kohn, 1977; Blau, 1982). Slaughter concludes from a

review of the literature on parental influences on education (1983) that parental values toward education in general, and achievement in school in particular, are important determinants of their offsprings' motivation to become knowledgeable and, thus perhaps the salient influence on scholastic attainment of young people as they mature. With more education, parents tend to be more vigilant of their children's academic achievement and more likely to consult their children's teachers and principals about their children's progress (Douglas, 1964). Better educated parents appear more attuned to the requirements of upward social mobility, particularly the importance of scholastic attainment, and they consciously prepare and encourage their children and monitor their progress in meeting these requirements (LeVine, 1973).

In addition to transmitting behavioral orientations to her children socially, the mother transmits half her genes to her offspring, and thus influences biologically the behavior of her children. Moreover, there is in the United States and other developed countries considerable assortative mating with respect to intellectual characteristics (Vandenberg, 1972). As a result, the attainments of the mother are substantially correlated with those of the father, and thus provide a convenient and often more accessible index of the father's contribution to both cultural and biological transmission of behavioral traits. For all these reasons, the mother's characteristics are among the best predictors of the characteristics of her offspring at maturity.

#### 6.1.3 Mother's Education and Socialization

Erickson (1968), having observed considerable differences in the childrearing practices of diverse cultures, concluded that there is a logic to socialization of children based on the group's understanding of what the child is to become and where he or she will find a livelihood. Parents' social values,

learned in the surroundings in which they grew up, and their concepts of competence based on personal experience or that of friends and relatives, will determine what they believe the child can achieve and how they may help him or her to realize that goal. It is not surprising, therefore, that educated middle-class mothers are able to induce skills and attitudes in their children that give them long-term advantages over lower-class children in competition for position and resources. We have commented on the fact that middle-class families are more attuned to the need for mobility and resourcefulness, and consciously set about preparing their children to meet these requirements.

Referring to the childrearing of university-educated American parents, Levine (1973) characterizes their practices as an attempt to "rationalize the process of socialization by consciously gearing both ends and means to information feedback from a changing environment" (p. 516). Levine suggests that educated parents are conscious of the fact that criteria for evaluating behavior change with each succeeding generation and that their children will be assessed by different standards than they have experienced. These parents therefore search for information on the emerging selective pressures, relying on childrearing experts as a resource, and self-consciously acting as mediators of environmental feedback to their children. In rationalizing the socialization process, highly educated parents are attempting to transmit their socioeconomic advantage to their children.

The results of the present study can be understood as an outcome of that process. The strong mother's education effects observed for performance in academic areas, especially in Mathematics Knowledge, Arithmetic Reasoning, General Science, Word Knowledge, and Paragraph Comprehension, indicate that more highly educated parents are rewarded with a high level of skill development in their children. Parents from groups that have not fully

participated in the American "mainstream," however, have less basis for knowing what competencies are or will be required for social mobility in their children's generation. Research also shows that mothers in the lower social strata are not aware that they can influence the development of their children's skills through specific activities (Tulkin and Kagan, 1972). Chan and Rueda (1979) conclude that social class differences in parenting behaviors arise from "inadequate access to knowledge and information concerning the requirements and techniques of socialization" (p. 425). These effects offer some explanation for the difference in response to mother's education seen in different sociocultural groups in Figure 6.1-3.

Why differential access to this type of information and its implementation exists is not clear. What is clear is that socialization strategies vary among sociocultural groups and that one of the consequences is intergenerational transmission of inequality. There is no question that various sociocultural groups show different levels of skill development. Nor is there any question about the probable consequences of differential skill development for vocational prospects. However, whether or not persons in low-scoring groups can be helped to develop greater levels of skill through providing more educational and vocational information to parents, and to mothers in particular, is an open question. We remain optimistic that low-achieving individuals in these groups can be helped to develop competitive skills through better counseling of parents and children, programs to broaden vocationally relevant experiences, and more instruction in basic skills. While some may argue that such efforts have been tried and have failed, they should be reminded that these earlier programs were generally poorly conceived, based on erroneous assumptions regarding the nature of the problem and the efficacy of different interventions, and hastily implemented (see Horowitz and Paden, 1973, for a critical review).

A direct and promising strategy would be greater exposure to a variety of occupations. Isolated, poor, minority status young people would profit from this increased exposure, especially if accompanied by clear discussions of the necessary skills for success in these occupations. It is now believed that one factor contributing to the lower academic achievement of poor and minority status young people is their failure to perceive the relevance of specific school subject matter for their later occupational prospects (Ogbu, 1978). Generally, teachers do not have the time, or do not take the time, to demonstrate the relevance of academic subject matter. Family members are less likely to be able to perform this function effectively because, with their limited educational and occupational experiences, they do not have the first-hand experience or specific knowledge to do so. While better vocational information and counseling may not eliminate variations in skill development, it seems a reasonable and efficient strategy to attenuate the differences.

## 6.2 Age and Education

The analyses discussed in Chapters 4, 5, and 6, do not include age as a background factor because the addition of another factor would have made the design unmanageably large and spread the data too sparsely over the smallest subclasses. Thus, the effects of highest grade completed are not distinguished from those of age in these discussions. Clearly, most subjects with some college will be older than those with less than 12 years completed and so on. The increase in test performance with increasing years of education seen in Chapters 5 and 6 could thus be equally well interpreted with respect to increasing age. To distinguish age from education effects, we must examine the test performance of persons of different ages who have the same number of years of schooling.

In this investigation we take advantage of the fact that mother's education interacts only in a minor way with other background factors. It justifies the collapsing of that classification so that age can be added. By ignoring mother's education, we can fit a model similar to that represented in Table 4.3-2, but also include age, age by sex, age by sociocultural group, age by highest grade completed, age by poverty, and age by sociocultural group by highest grade completed. The multivariate tests indicate that all but the triple interaction of age, sociocultural group, and highest grade completed are significant. We therefore fitted a model including all age-related terms except the latter.

For present purposes, we are primarily interested in the age by education interaction. Of the six possible contrasts for interaction of the three age groups and four education levels, only four can be estimated. The two remaining contrasts are confounded because there are no subjects with one year of college in the 15-17 age group, and few who have completed high school. When plotting the effects of age by highest grade completed, as in Figure 6.2, we therefore omit in the 15-17 year age group the points for 12 and 13+ grades completed.

We are justified in examining the age by grade interactions if these two factors do not appear in any significant three-factor interaction. Inasmuch as the most likely candidate--the age by grade by sociocultural group interaction--was not found to be significant, we will assume the triple interactions of age and grade with sex, region and poverty are also nonsignificant. We therefore proceed to interpret the estimated age effects in subjects with similar amounts of education as indexed by highest grade completed. Effects of all other background factors are eliminated from these estimates.

Age and grade effects on test performance are difficult to interpret because at least four distinct influences are at work to increase or decrease scores:

1. Average age at grade. It is well known to educational researchers that children who are more than six months younger than the average age for their grade in school have generally higher scholastic ability than their classmates. This occurs because such pupils include the highly precocious who have been advanced over one or more grades. Generally, these children are not advanced so far that they fall below the average of the class in achievement; in fact, they are usually found near the top of their class in spite of their younger age. For the opposite reason, children who are more than six months older than the average age for their grade in school tend to be below average in ability; they include children who have been held back until they can meet minimal standards for their grade. Because of the age-at-grade effect, we expect deficit in test performance when an older group is compared with a group that is at the expected age for grade.
2. Lack of practice and forgetting. When people of any age leave off formal schooling, they may begin to forget, through lack of practice and disuse, the more academic content of what they were taught. Arithmetic and mathematical skills are particularly susceptible to these effects because their practice is left mostly to specialists. Most students retested on this type of material some years after leaving school will tend to show decreased scores.
3. Experiential learning. The converse of 2 above, some skills of a practical nature tend to improve with experience, at least into middle age. Word Knowledge, for example, tends to increase somewhat through the third and fourth decade of life as the accumulative result of contact with language.
4. Cohort effects. Changing curricular emphasis and instructional practice may be reflected in differences between average test scores of cohorts who passed through the school systems at different times. The well-publicized decline in nationwide average Scholastic Aptitude Test Scores is believed to be due in part to changes in the curriculum and relaxing of academic standards in the high schools during the 1960s and '70s. These effects can appear as age differences when cross-sectional data are analyzed. But they are really cohort differences that have nothing to do with developmental changes within persons.

Some or all of the above influences may be seen in the results plotted in Figure 6.2. It is apparent, however, that when these influences are at work, they are acting in much different combinations in the various ASVAB tests. Among respondents with the same amount of schooling, the Electronics Information and Auto and Shop Information scores appear to be generally benefitting from the experience that comes with age. As we observed earlier, this is the kind of knowledge that is acquired mostly outside of school, so much so that persons with only an 8th grade education can increase their scores over the years from 18 to 22 just as much as those in college. A similar but weaker effect is seen in Mechanical Comprehension.

One reason that experiential learning effects can readily be seen in these types of tests is that they are not criteria for scholastic advancement or holding back. They are therefore not subject to the age-at-grade phenomenon that works in the opposite direction (i.e., produces higher scores in younger persons with the same amount of formal education). This permits gains with increasing age to be clearly seen.

Word Knowledge, General Science and possibly Paragraph Comprehension also measure knowledge that would be expected to accrue with experience in this age range. But these tests are closely related to criteria for educational attainment, and so would be subject to the age-at-grade phenomenon. The result is consistent with what we see in the plots for these tests: rather small changes, some positive and some negative, between ages within education levels.

Mathematics Knowledge, Arithmetic Reasoning, possibly Coding Speed, and most prominently, Numerical Operations, appear to combine the age-at-grade phenomena with the deleterious effects of disuse and forgetting. The older groups at the same education levels perform more poorly, in some cases much



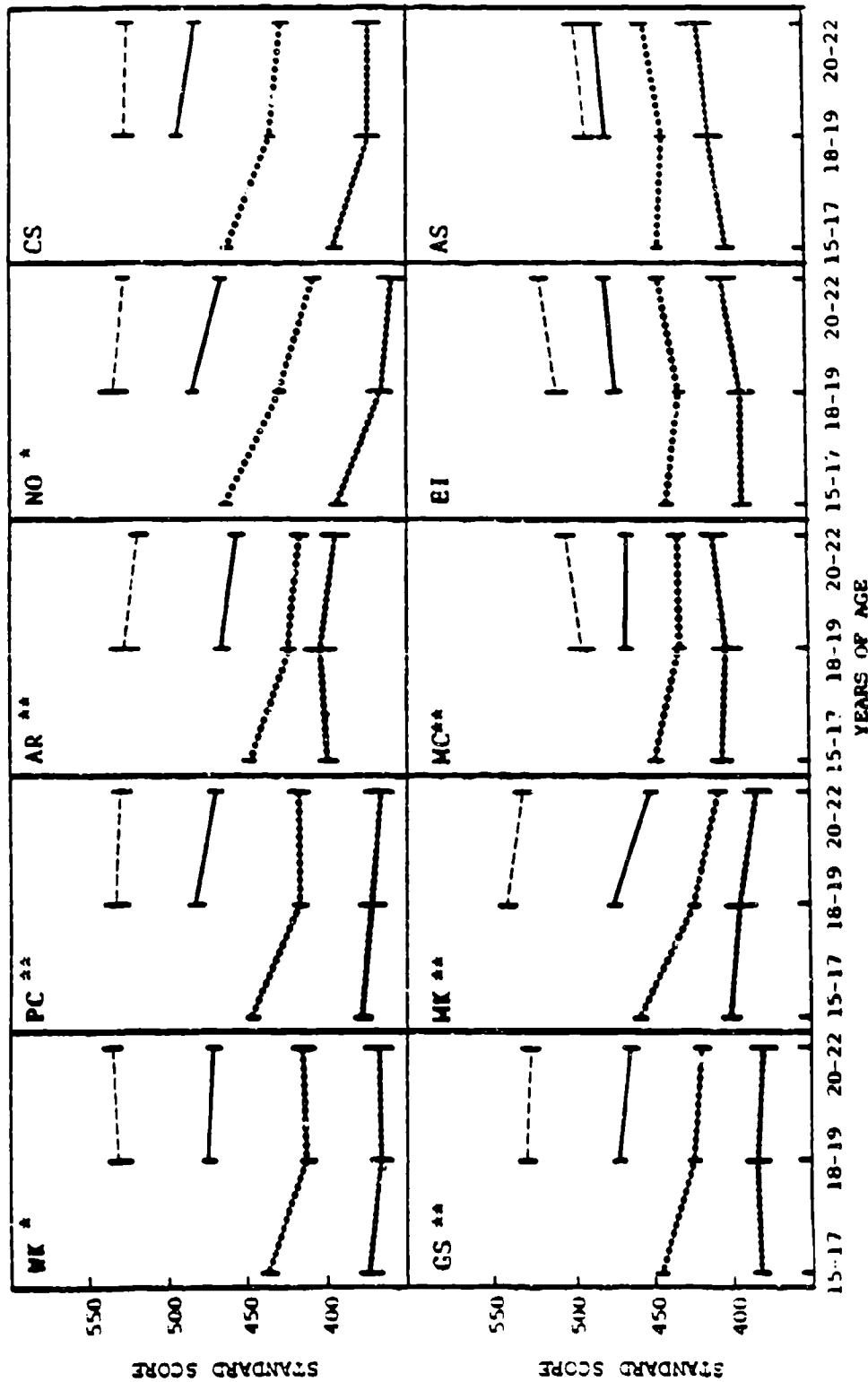


Figure 6.2 Interaction of age and education (highest grade completed).

Grades 0-8 .....  
 Grades 9-11 .....  
 Grade 12 .....  
 Grade 13 or more .....  
 \*F>1.25 F(.05)  
 \*\*F>1.25 F(.01)

more poorly, than the younger age groups. The contrast with the Electronics Information and Auto and Shop tests in this regard suggests the deterioration of skill rather than continued learning once formal education is terminated. A more optimistic interpretation would be that in these quantitative areas we are seeing cohort effects resulting from improved instruction in mathematics. If the younger groups received better instruction in quantitative skills they might score better on these particular tests. For this to happen in the Profile sample would, however, require a nationwide change of instructional practice in a relatively short time--an unlikely possibility. Lacking any independent evidence for such a change, we might explain these effects more reasonably as the result of knowledge and skills in the numerical and mathematical area not being maintained as well as those in the more verbal and practical topics. This is less true among college level people, and certainly would not be true of young people specializing in these areas, but such people are relatively few and do not much influence the broadly-based sample in the Profile study.

The more general conclusion that must be drawn from the mean scores displayed in Figure 6.2 is that, for the most part, test performance tends to be typical of the highest grade completed. With possible notable exceptions in individual cases, most of those young people who drop out of school with not more than eight grades completed, or without graduating from high school, have little opportunity or inclination to improve their abilities in areas represented by the more academic of the ASVAB tests. Even their improvement in the practical areas is small relative to the large differences between education levels.

The deleterious effect on vocational opportunities of lower performance levels of persons who drop out of school will be considerable and will

ameliorate little as they grow older. The broader educational implication of the age-by-grade effects would seem to be that efforts toward more complete schooling of all young people who are to any reasonable degree capable of it will benefit both them and the society at large.

### 6.3 Summary and Conclusions

This chapter deals with two variables influencing vocational test performance that are related by their direct role in the personal development of young people. One is the years of education of the mother; the other is the number of years of schooling that a person has completed by a given age.

#### 6.3.1 Mother's Education

Increasing level of mother's education--for grade school only, some high school, completed high school, some college, to completed college--is directly and strongly related to higher scores on all tests in the ASVAB battery. The relationship is stronger for the tests that are more academically oriented, namely, Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, Numerical Operations, General Science and Mathematics Knowledge. It is weaker for the tests that assess knowledge acquired experientially, namely Mechanical Comprehension, Electronics Information, and Auto and Shop Information, and also for Coding Speed, which is a performance task. The implication is that the mother's educational values, if not her direct instruction, press her children for achievement in areas that are part of and identified with school learning. Other types of attainment, while benefiting from better home resources (and possibly heritable potential) contributed by a better educated mother, are less directly affected and do not increase as much with increasing years of education of the mother.

Mother's education shows only limited and very minor interactions with other background factors. In very technical topics such as Electronics Information, there is a slight but significant tendency for sons to benefit more than daughters from increased mother's education. Presumably this is due to sons making better use of mother's contribution to home resources in this male-stereotyped area.

A somewhat stronger interaction appears between sociocultural group and mother's education for two of the tests, namely, General Science and Auto-Shop Information. It is due entirely to improved performance of offspring of Hispanic mothers with education beyond high school. Studies of course participation in higher education suggest that Hispanics who are oriented toward higher education favor technical programs. Those with less academic interests, but whose family backgrounds direct them toward more economically rewarding occupations, frequently choose mechanical and construction trades. These tendencies appear to be more marked in Hispanics than in Whites or Blacks, and thus appear to explain the interaction.

The fact that these interactions are small relative to the overall effect of mother's education, and all other interaction of background factors with mother's education are statistically insignificant, indicates that variables that influence achievement through the mother's contribution to her offspring are largely independent of the broad social and cultural effects on attainment that we have examined in Chapter 5. They are for the most part also independent of the influences of sex of offspring that we will consider in Chapter 7.

Studies in the literature that have inquired into the role of the mother in shaping the educational and vocational achievement of her children emphasize the guidance she can provide to where the best opportunities lie in

these areas. Better educated mothers are much better informed in this respect than the poorer educated, and their children benefit accordingly. Programs that provide greater counseling and experiences with vocations, both to parents and offspring, have promise for extending educational and vocational opportunities to all segments of society.

#### 6.3.2 Age and Education

Examining mean scores of persons in the age range 15 to 17 who have completed up to 8 grades compared to those who have some high school, and those in the 18 through 19, and 20 through 22, age groups who have up to 8 grades, some high school, completed high school or some college, we find evidence that performance on some of the ASVAB tests improves with age independent of education and on others deteriorates. As we might expect, it is the school intensive subjects, Mathematics Knowledge, Arithmetic Reasoning, and, most prominently, Numerical Operations, in which the older respondents at the same education levels perform more poorly. Perhaps the most plausible explanation is simple forgetting of facts and loss of skills due to disuse.

Conversely, the knowledge that is acquired and maintained by experience outside of school, represented here by Electronics Information and Auto and Shop Information, improves with age among respondents with the same education. Areas represented by the remaining tests are intermediate between these extremes.

These age by education interactions are not, however, the dominant impression conveyed by these comparisons of respondents classified by age and highest grade completed. The more general conclusion is that, regardless of age, vocational test performance tends to remain near the level of the highest grade completed. In particular, failure to complete high school augers poorly for meeting vocational test standards at a later time through informal learning and experience.

## CHAPTER 7

### SEX EFFECTS

As increasing numbers of women enter the labor market, sexual inequalities in educational and occupational attainment have become a cause for public concern and debate. Census figures confirm that females are underrepresented in the more prestigious and higher paying occupations such as engineering, science, mathematics, business management, and the skilled trades. Why these disparities should exist, and the extent to which they are conditioned by biological characteristics as opposed to sociocultural sex role expectations and proscriptions, are questions that have stimulated much research and controversy.

From a comprehensive review of research on sex differences in behavior, Maccoby and Jacklin (1974) conclude that certain cognitive differences between males and females are clearly established: beginning in early adolescence and increasing through high school, girls have greater verbal ability than boys; beginning in adolescence and continuing to maturity, boys increasingly excel in visio-spatial ability and in mathematical ability.

Jacklin (1979), however, has warned that these earlier conclusions must be viewed with caution because of the limitations of the data from which they derive: nearly all of these findings are based on samples of White middle-class subjects. She considers the apparent sex-related differences in cognitive abilities, though potentially significant for occupational choices, to be only tentatively established and to require broader verification.

In large measure the present study provides precisely this verification. The Profile data assess sex-related differences in cognitive abilities in the larger socioeconomic and cultural context of this country. These data

are, of course, descriptive and cannot be used to address all issues concerning sex differences. In particular, they do not speak directly to the question of causes. But combined with an extensive research literature, they provide a background for theorizing about the origins of sex-related differences in performance observed in this sample.

To put the sex differences in their proper perspective, we must take note of the fact that in the entire sample, subject's education has the strongest relationship to level of performance for all the subtests. Sociocultural group membership has the next strongest relationship, mother's education is third, and sex and region are fourth and fifth respectively. Thus, sex is by no means the most important predictor of performance on the ASVAB. In addition, sex classification interacts with education and with sociocultural group and economic status. Therefore, while main effects of sex are observed (females generally perform significantly better than males in Paragraph Comprehension, Numerical Operations, and Coding Speed; whereas males perform better than females in Arithmetic Reasoning, Auto and Shop Information, Mechanical Comprehension, Mathematics Knowledge, and Electronics Information), these effects are best understood in the context of the interactions between sex and other background variables.

The picture is simplified by the fact that the stronger interactions do not extend beyond sex and education, and sex, sociocultural group, and economic status. The interactions of sex and age, sex and region, or sex and mother's education, are minor in comparison.

#### 7.1 Sex by Education Effects

The interaction between sex and highest grade completed is displayed in Figures 7.1a and b. The performance of males and females with 0 to 8,

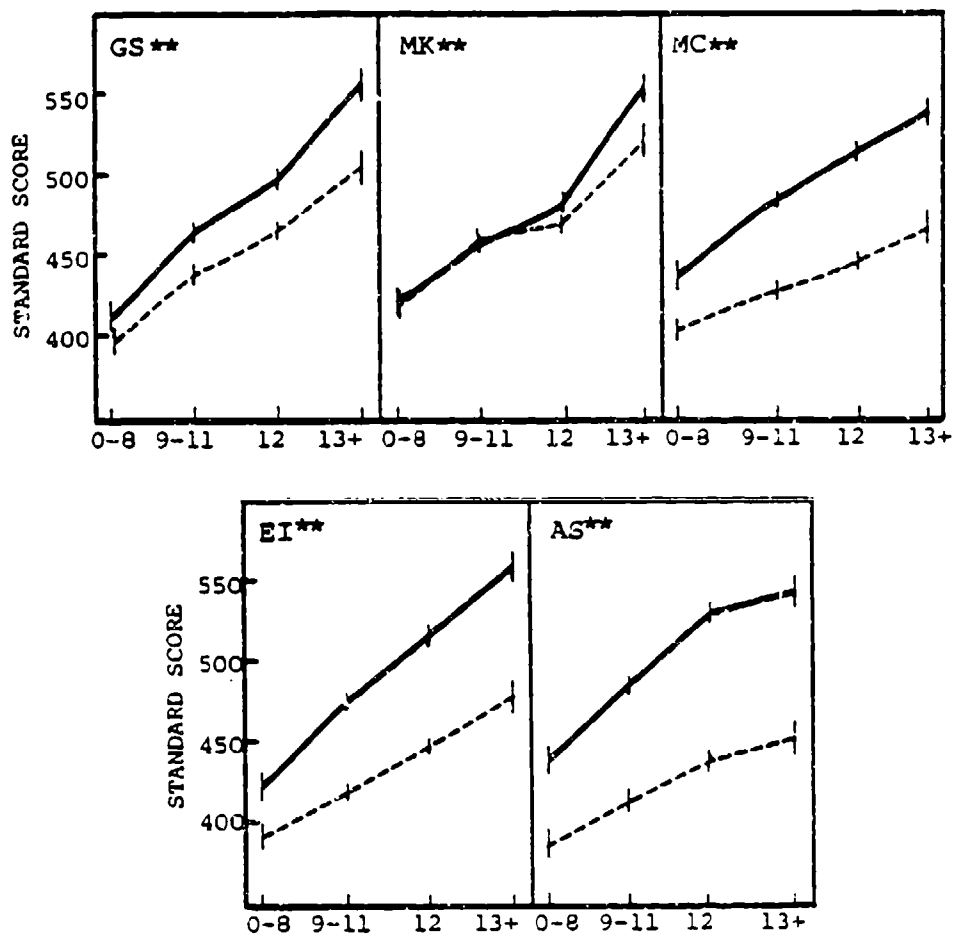


Figure 7.1a Interaction of sex and education  
(highest grade completed).

Male——Female-----

\*\*F > F(.01)



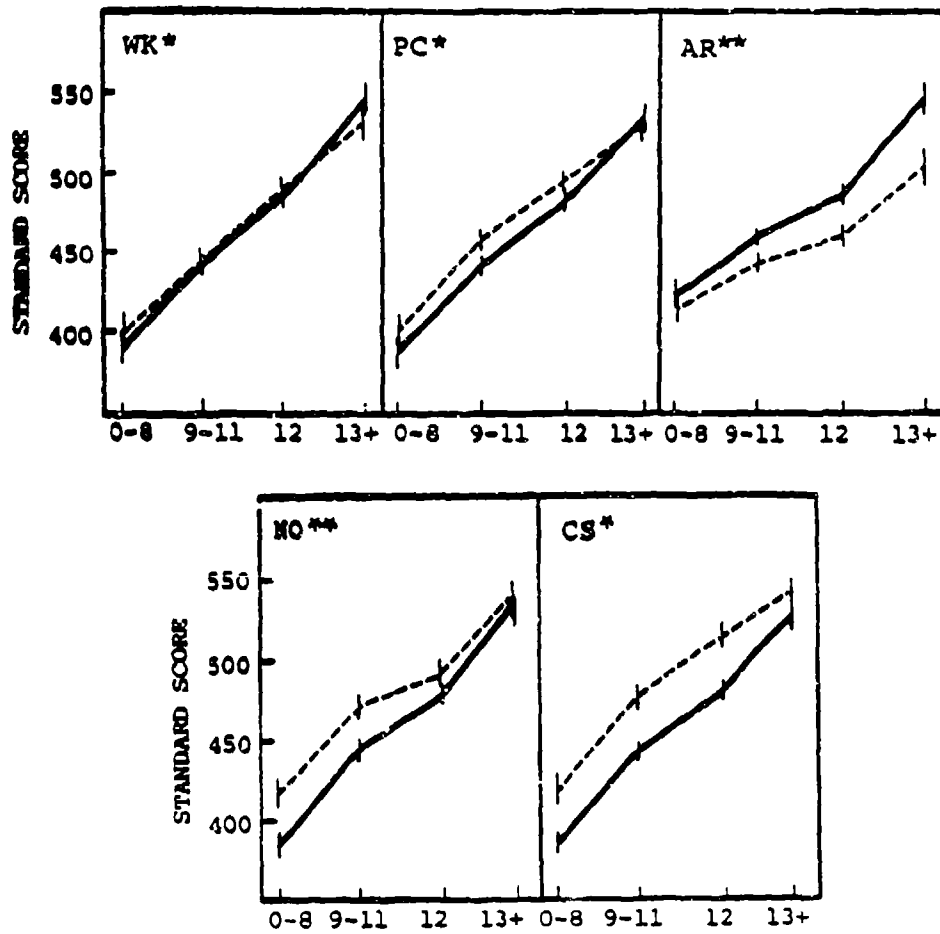


Figure 7.1b Interaction of sex and education  
(highest grade completed).  
Male——— Female-----

\* $F > 1.25F(.05)$   
\*\* $F > 1.25F(.01)$

9-11, 12, and 13 or more years of education is shown in terms of the standard scores of the total probability sample for each subtest (mean=500, standard deviation=100). These graphs reveal the various patterns of sexual differentiation in test performance by educational level discussed in this section.

Specialized Knowledge. Sex differences in achievement increase as educational attainment increases for those tests that assess technical knowledge--namely, General Science, Auto and Shop Information, Mechanical Comprehension, and Electronics Information. These areas of knowledge could be described as traditional male domains where sex differences would be expected.

In General Science, the males in the sample with 0 to 8 years of education score only slightly higher than females. However, with some high school and high school completion, the difference in the scores of males and females increases to about 35 points, and the largest difference between males and females in General Science occurs at the 13+ level (about 50 points).

A pattern of increasing difference between males and females with education also is seen for Mechanical Comprehension. At the lowest educational level, the difference in favor of males is 30 points; with some college, the males outperform females by more than 70 points.

The picture is similar for Electronics Information, but in both sexes the mean scores are somewhat more strongly related to education than is the case for Mechanical Comprehension scores. Except for the greater sex difference, the Electronics Information plots look more like those of General Science. We take this as an indication that Electronics Information and General Science depend more on knowledge of technical facts or principles gained through formal study, while Mechanical Comprehension contains elements that can be learned by practical experience of mechanical relationships. Women are less in contact with, or less perceptive of, the latter and for Mechanical Comprehension show the least gains with education.

While males achieved higher scores than females on each of these tests at each level of education, the fact that the performance of both males and females increases with education deserves emphasis. In both groups, more education appears to enhance performance, but in General Science, Mechanical Comprehension, and Electronics Information, education at high school and beyond is associated with better performance among males than among females.

The possible origins of these sex differences are discussed in section 7.3. But one conclusion is obvious: large sex differences occur in those areas that traditionally have been preempted by males. The implications of this fact must be explored.

The Auto and Shop Information test shows the greatest differences between the sex groups at each education level. Males at the lowest educational level out-perform females with similar education by 70 points. Males who have completed high school or have some college are a full standard deviation (100 points) above females at each of these educational levels. As Figure 7.1a shows, male performance on this test exhibits a much greater benefit from education through completion of high school than does female performance. Education beyond high school, however, results in only a small increase in the mean scores for males, while scores for females rise steadily over all education levels.

These results have a plausible explanation in the observation that auto and shop information is highly dependent upon experience. Males at present are much more likely to have experiences in this domain in their jobs and in their hobbies than are females. They are also more likely to participate in the largely male stereotyped auto and shop school electives. It is thus not surprising that females' performance on this test is considerably lower than that of males at every educational level.

In addition, correct responses to the auto and shop items require less knowledge of principles than do Mechanical Comprehension and Electronics Information, and, as we see, performance does not change with college education to the extent that is seen in Mechanical Comprehension and Electronics Information among both males and females. By the same token, the greater change in female performance between high school and college probably reflects the greater tendency of females to develop this information from reading or vicarious experience rather than hands-on experience. Hyman, Wright, and Reed (1975) have shown that as education increases, people's seeking of information in general also increases. Having learned less of this material in high school (as Figure 7.1a suggests), better educated women have an opportunity to increase their knowledge in this area by reading, better understanding of mechanical principles, and general experience. Finally, the result could also be explained by women who elect to attend college having wider interests extending even to such traditionally male domains as auto and shop.

Related to the sex difference in the other specialized knowledge tests, the current literature substantiates the common observation that in high school and college a higher proportion of males than females enroll in science courses such as physics and chemistry, and in shop and mechanical drawing courses (Cook and Alexander, 1981; Fennema, 1981; Sells, 1980; Erlick and Lebold, 1977). Males also show greater interest in science in the early school grades (Fox and Cohn, 1980), and their out-of-school activities more often involve work with power tools and reading magazines such as Popular Science and Popular Mechanics (Erlick and Lebold, 1977). Females generally avoid these knowledge domains in high school and college because they do not perceive vocational opportunities in fields where this information is required and/or they do not aspire to careers in these fields. The exception is

biology, which is more often taken by women (Marrett, 1981). As a result, the General Science test, about half of which is devoted to biology, shows smaller sex differences than Electronics Information. Generally, the pattern of group means on the ASVAB specialized knowledge tests is consistent with the different patterns of high school and college study typical of males and females.

Verbal Attainment. The second clear pattern in the sex group means involves the Word Knowledge and Paragraph Comprehension tests. In contrast with the increasing sexual differentiation with education seen in the tests of specialized knowledge, differences between the mean scores of males and females are small for the subtests that assess verbal ability. As Figure 7.1b shows, females show a very small advantage over males in Word Knowledge at the lowest educational level. This advantage becomes smaller at each higher educational level, and ultimately gives way to an advantage for males at the 13+ level. The differences are very small, however, and the statistical significance of Sex by Highest Grade Completed interaction in Word Knowledge is marginal (see Table 4.3-1).

A pattern of increasing similarity in the performance of males and females is also seen in the Paragraph Comprehension test. However, females show a larger advantage in this area at the lower educational levels than they do in Word Knowledge, and their advantage is nearly constant to grade 12. At the highest educational level, the average scores of males and females on Paragraph Comprehension are essentially the same.

These patterns of sex differences in verbal ability generally confirm previous research findings in this area. As noted earlier, Maccoby and Jacklin (1974) conclude from their review of the literature that sex differences in verbal ability occur more consistently in early adolescence or at the time of entry into secondary school. The differences between the sexes are

found primarily in verbal fluency and reading comprehension, where girls are higher achievers.<sup>1</sup> The general advantage of females on the Paragraph Comprehension test (which essentially assesses reading ability) at each educational level except the highest is therefore not surprising. The convergence of the scores of males and females at the highest educational level can be attributed to the greater probability that both males and females who pursue education beyond the high school level reach a mastery level of reading skill.

The Word Knowledge test shows a very slight trend of difference between males and females with increasing education to the point of a slight male advantage at the highest educational level. This would not be expected on the basis of existing literature. Generally, previous research does not indicate any consistent sex difference in tests of word knowledge (Maccoby and Jacklin, 1974). In recent years, however, men have been attaining a slightly higher average score than women on the Scholastic Aptitude Test (SAT) verbal scale. The difference in favor of males grew from 3 points (on a scale from about 200 to 800 points) in 1973 to 12 points in 1981 (College Board, 1981). At the same time the quantitative score difference favoring males grew from 42 to 49 points. These changes may be due to the larger, and hence less select, group of women who now take the SAT tests. In 1981, 515,708 women and 478,625 men took the tests, and the relative number of women taking the tests has been increasing.

-----  
<sup>1</sup>Verbal fluency is the ability to produce a quantity of verbal output very quickly. In the Word Fluency test of Thurstone and Thurstone (1947), for example, subjects are required to write in five minutes as many different words as possible beginning with the letter "s." From 8 or 9 years of age onward this test shows large sex differences favoring females.

It is also possible that males who remain in school are exposed to a greater variety of language contexts. In addition to fulfilling course-distribution requirements for high school and college graduation, they are more likely to enroll in elective and major courses not typically taken by females, such as science and technical courses. Additionally, differences in hobbies and leisure time engaged in by males and females may result in males' exposure to more diverse language domains and vocabulary than females.

Although these observations are somewhat speculative, the fact remains that in the Profile data males with more than a high school education show slightly greater Word Knowledge scores than do females with comparable education. For both males and females, the higher the educational attainment, the greater the Word Knowledge and Paragraph Comprehension, but the educational benefit to males' performance in these areas is greater, particularly at the highest educational level.

Fluent Production. A third pattern of performance difference between male and female respondents emerges for the two speeded tests--Numerical Operations and Coding Speed. Figure 7.1b shows that, while the performance of both males and females on both of these tests increases as educational attainment increases, Coding Speed shows a consistent large difference, favoring females, between the two groups across educational levels. A similar pattern is seen in Numerical Operations through grade 12. With some college, however, the scores of males and females are nearly identical.

As we discuss below, the difference between the two sex groups in Coding Speed, and, to a lesser extent, Numerical Operations suggests the working of the well-known mental test factor of fluent-production on which females typically excel. Also, female students more often than males enroll in secondary school courses that develop fluency skills such as typing and

stenography, (Erlick and Lebold, 1977; Cook and Alexander, 1981). But why educational attainment should relate so strongly to performance on these simple tasks is not as obvious. It may be that with increased schooling people develop more efficient test-taking skills, particularly in symbol recognition and short-term memory. Motivational factors and persistence factors may be implicated as well. Since more highly motivated young people are likely to seek more education, such people are also more practiced in the rote manipulation of conventional symbols than are people with less education.

Quantitative Attainment. In contrast with those for Auto and Shop Information, scores on the Arithmetic Reasoning and Mathematics Knowledge tests are highly associated with education, especially so when subjects with only high school education are compared with those in college. At the lowest educational levels (0-8 and 9-11), Figures 7.1a and b show no differences between male and female achievement on Mathematics Knowledge and relatively small differences in Arithmetic Reasoning. The amount of increase in male and female scores in the lower educational levels is identical. The performance of the two sexes begins to diverge for subjects with 12 years of education, though the difference in favor of males at this educational level is small, about 12 points. At the highest educational level, the advantage of males over females increases to about 30 points. For Arithmetic Reasoning, on the other hand, a sex difference favoring males is readily apparent among subjects who have completed high school and increases to about 50 points for those with some college education.

In addition to the sex differences in the patterns of education effects in Mathematics Knowledge and Arithmetic Reasoning, the male advantage over females in Mathematics Knowledge is not as large as that seen for Arithmetic Reasoning. We attribute these differences to the fact that the



Mathematics Knowledge test requires the literal application of mathematical rules (see the sample items from this test in Chapter 3), whereas the Arithmetic Reasoning test requires a restructuring of the verbally presented material before an algebraic rule can be applied (Fennema, 1981).

There is no reason to conclude that males and females who have had similar exposure to rules involved in Mathematics Knowledge would differ significantly in their ability to apply them. The equal performance of the two sexes at the lowest education levels, and the higher levels of performance of both males and females with increasing education, support this point. The divergence of males and females in mathematical knowledge at the higher educational levels can be explained by sex differences in rule acquisition due to different participation in higher level mathematics courses. In contrast, the form of problem restructuring required in Arithmetic Reasoning is aided by non-verbal thought (i.e., by mental manipulation of images requiring spatial visualizing ability), and the well-documented sex-differences in this ability may contribute to the different pattern of sex effects in these two types of mathematics tests. (We review this evidence in Section 7.4.)

## 7.2 Sex by Sociocultural Group by Economic Status

Performances of poor and nonpoor males and females in the White, Hispanic, and Black subsamples are presented in Figures 7.2a and b for each of the subtests in terms of the standard scores of the probability sample. As can be seen, there are considerable differences in achievement between the income and sociocultural groups on the various tests. We have discussed these overall group differences in Chapter 5. In addition, interactions between sex, sociocultural group, and economic status appear as nonparallelisms in Figures 7.2a and b. These interactions are moderate in size but clearly significant.

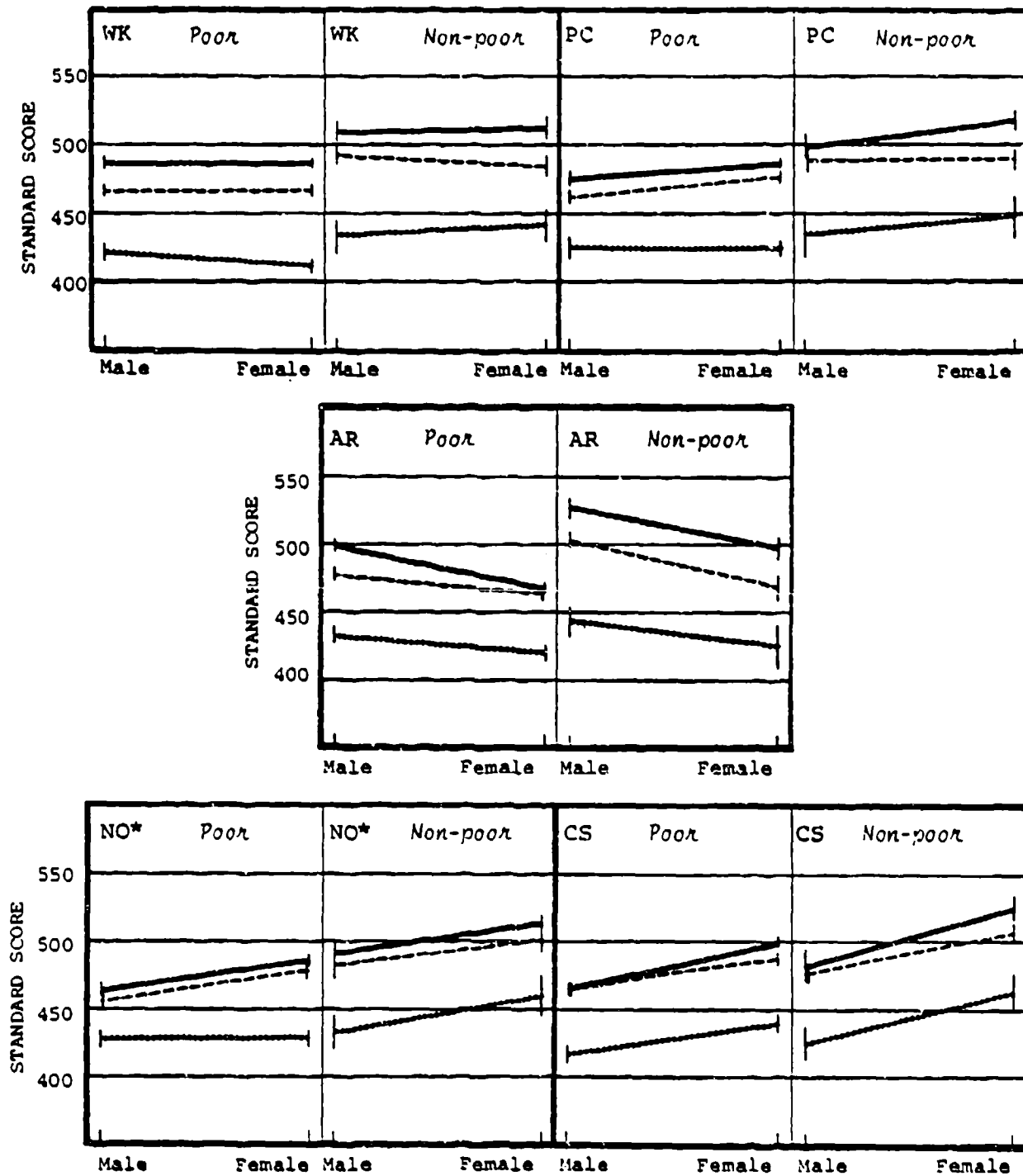


Figure 7.2a Interaction of sex, sociocultural group, and economic status.

White——— Hispanic----- Black.....

\*F > 1.25F(.05)

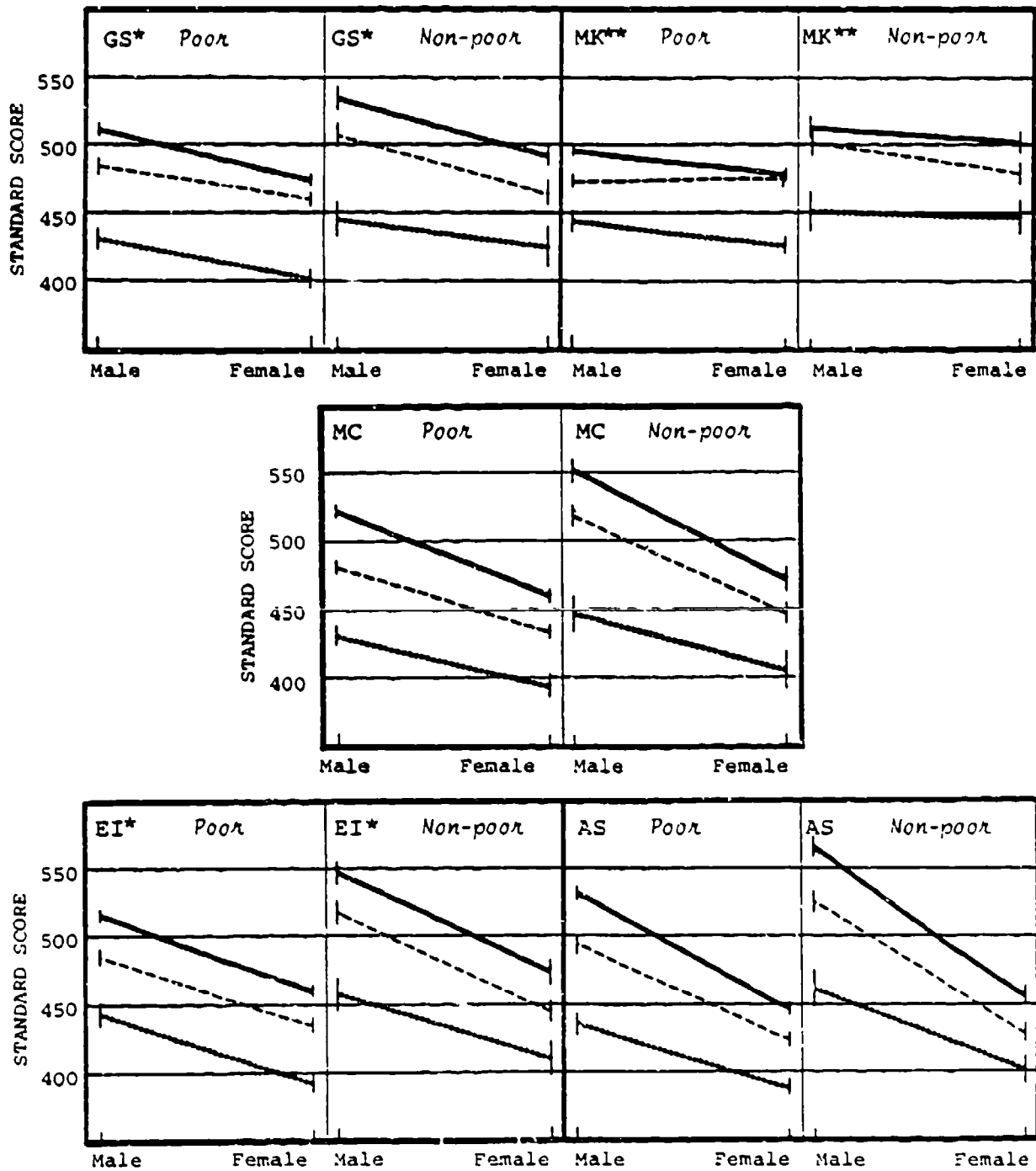


Figure 7.2b Interaction of sex, sociocultural group, and economic status.

White ——— Hispanic - - - - - Black . . . . .

\*F > 1.25F(.05) \*\*F > 1.25F(.01)

Specialized Knowledge. On the tests that generally show male advantage (i.e., General Science, Mechanical Comprehension, Electronics Information, and Auto and Shop Information), the size of the sex difference in performance is affected by both sociocultural group membership and income level. The largest sex differences occur in the nonpoor White and nonpoor Hispanic groups. The males in these groups show almost identical patterns of relative advantage over their female counterparts. Likewise, poor Whites and poor Hispanics show a similar pattern of male advantage, though the male advantage in these groups is smaller than that of their higher income peers. Blacks generally show less sexual differentiation of performance and smaller effects of income on the size of the male advantage.

The General Science test gives strong evidence of these interactions. The males in both the nonpoor White and nonpoor Hispanic groups score about 40 points higher than the females in these groups. Poor White males exceed the performance of their female counterparts in General Science by about 35 points, which is only slightly less than the sex difference seen in the nonpoor White subsample. The difference in the average scores of males and females in the poor Hispanic and poor Black subsamples is similar to but smaller than that seen among poor Whites, amounting to about 25 points. The most salient instance of a reduced sex effect in the specialized knowledge tests occurs, however, among nonpoor Blacks in General Science. In these groups, the sex difference in General Science performance is only about 12 points in favor of males. The fact that the Black culture, more so than that of Hispanics and Whites, has traditionally encouraged females to seek personal and familial advancement through education (Slaughter, 1972; Edwards, 1979) must be contributing to the smaller sex difference seen for nonpoor Blacks in achievement in General Science. Allen's (1980) analysis of the National Longitudinal Survey

(NLS) data for the High School Class of 1972 provides evidence of the high educational expectation of Black females. This investigation revealed that while the educational aspirations of Whites and Blacks are comparable, the percentages of Black females and White males with goals of higher education are slightly greater than was true of their respective male and female counterparts. Since this is the case, we would expect that, relative to Whites, more Black females enroll in college preparatory programs during high school and participate in more science courses. It is therefore reasonable to expect less sex differentiation among Blacks than among Whites in their knowledge of General Science.

This inference is supported by Marrett's (1982) recently conducted survey of 42 high schools in 36 school districts, which revealed that the science course participation of Black females and males is roughly equal, even in specialized courses such as chemistry and physics. This is not true of Whites. However, Marrett also reports that the proportion of Blacks of either sex who enroll in these science courses in high school is considerably smaller than that seen for White males or females.

Since Black females generally show higher educational expectations relative to Black males, regardless of income class membership, and a higher proportion of Black females than males in both low and high socioeconomic categories enter college (Thomas et al., 1979), it is surprising that the sex effect seen in the poor Black sample is not more similar to that seen for higher income Blacks in General Science. The fact that the male advantage in the poor subsample is greater than that seen in the nonpoor category suggests that the science course participation of the sexes is more similar among higher income Blacks than poor Blacks. This may be due to poor Black females receiving less adequate educational guidance during the high school years than

nonpoor Black females. Turner's observations from her 1982 investigation of factors that influence women's science career interests suggest that the educational and career guidance available to Black high school students may be inadequate, especially in portions of the southeast region. Turner notes (personal communication) that about 35 percent of the high school males and 45 percent of the females identified by school counselors as probable science majors in college were not enrolled in college preparatory programs of study. While this analysis did not examine the proportion of poor and nonpoor Black high school students whose programs did not fit their educational and career ambitions, it would not be surprising to find this situation occurring more frequently for poor Black students than for the nonpoor.

Fluent Production. In the tests where females generally show higher average performance than males (i.e., Coding Speed and Numerical Operations), only the Numerical Operations test provides evidence of significant interaction with economic status and sociocultural group. The interaction is due almost entirely to the fact that poor Black males are essentially equal in performance to their female counterparts in Numerical Operations. It is possible that the smaller sex difference in the poor Black group is due to females performing more poorly rather than males performing better. An explanation often advanced for females' higher performance in fluency tests is that females conform more to the demand for speed in these rather onerous tasks than males. But conformity to social expectations probably increases with socioeconomic status, so its effect would be less among the poor Black females.

Quantitative Attainment. Generally, males in each of the sociocultural groups show higher average scores in Arithmetic Reasoning than their female counterparts, regardless of economic classification. However, the

amount of difference seen in the scores of the sexes in the various categories of young people is not the same across groups. The smallest sex difference in Arithmetic Reasoning is seen in the poor Black and poor Hispanic subsamples, amounting to about 10 points in favor of males. The advantage of nonpoor Black males is somewhat higher, about 20 points, but less than the 35 to 40 point difference seen in both income classifications of Whites and in the nonpoor Hispanic subsample. Again we speculate that cultural differences, mediated by income level, contribute to the differences between groups in the amount of sexual differentiation in performance seen in Arithmetic Reasoning.

In view of the distribution of the different Hispanic groups in the various regions of the country (Brown et al., 1980), the fact that the highest proportion of poor Hispanics occurs in the West (see Chapter 2, Table 2.5-1) indicates that Mexican-Americans are substantially represented in the poor Hispanic subsample. Since this is the case, the literature indicating cognitive style differences between Mexican-Americans and Anglo-Americans suggests a possible explanation of the reduced sex difference seen for poor Hispanics in Arithmetic Reasoning. These reports suggest that Mexican-Americans who experience more traditional socialization patterns (which is most likely to occur among the poor) differ from Anglo-Americans in the construct "field-dependence/independence." Traditionally socialized Mexican-Americans tend to be less field-independent (i.e., perceptually capable of overcoming an embedded context) than Anglo-Americans (see Kagitcibasi and Burriel, 1977, for a critical review), and, Mexican-American females are more field-dependent than their male counterparts. Ramirez (1973) theorizes that culturally conditioned differences in this cognitive style between Anglo-Americans and Mexican-Americans contribute to the latter group's generally depressed scholastic achievement because the American school system assumes a more field-independent style in the formulation

of instructional techniques and the development of curriculum materials. A possible link between cognitive style and mathematics achievement, in particular, is suggested by the work of Vaidya and Chansky (1980), who investigated the relationship between relative field-dependence/independence and mathematics achievement. These investigators report that field-independent pupils exceed their field-dependent peers in mathematics achievement, and hypothesize that the relationship exists because "analytic ability required in field-independence and mathematics achievement tests involves perceptual disembedding and developing problem-solving strategies that depend on reorganizing and restructuring the information" (p. 326). A recent large-scale investigation by Gemmill et al. (1982) of factors that influence the mathematics course participation of highly able Mexican-American adolescents confirms the finding that relative field-dependence/independence as measured by the Group Embedded Figures Test (GEFT) is associated with mathematics achievement, and that Mexican-American youth score more in the field-dependent range than is true for Anglo-Americans.

But care must be taken in interpreting these findings. As Ramirez (1973) has suggested, the significant variable for mathematics achievement may not be cognitive style per se but rather how young people who differ in cognitive style are instructed in mathematics and other subject matter. In other words, the nature of the apparent relationship between culture, cognitive style, and mathematics achievement may not be straightforward.

The current literature provides clear evidence that minorities do not enroll in high school mathematics courses or pursue mathematics-related careers in the same proportions as their White counterparts (Matthews, 1983; National Science Foundation, 1980; Gemmil et al., 1982). A major finding presented by Marrett (1982) from her large-scale survey of minority females in high school mathematics and science courses is that there are far greater



sociocultural than sex differences in mathematics participation. Although Black students represented over half of the sample, Marrett reports that they were only one-third of the students enrolled in algebra II, and one-fourth of the students enrolled in calculus. The Black students in this sample were concentrated in lower level mathematics courses, that is, 48.9 percent of the Black high school students sampled were enrolled in general mathematics courses such as technical math, consumer math, business math, and other introductory level courses. In contrast, only 25.8 percent of the White high school students were enrolled in similar lower-level courses. Marrett also reports that a slightly higher proportion of the Black females were enrolled in algebra II, calculus and other advanced mathematics courses than Black males, though the proportion of Blacks of either sex enrolled in such courses was considerably lower than that seen for White males or females. One important consequence of lower levels of preparation may be reduced opportunity for sex differences to assert themselves. If so, the lower enrollment of Blacks in advanced mathematics courses generally, would, in large part, account for the lower sex difference in Arithmetic Reasoning seen for poor Blacks, and to a lesser extent, nonpoor Blacks.

The same may be true for poor Hispanics. The fact that Hispanics are considerably underrepresented in mathematics-related professions (see McCorquodale, 1980 for a review), and underrepresented also in science and mathematics majors in college (Astin, 1982), suggests that there may be little differentiation in the secondary school mathematics preparation of males and females in this group. Although McCorquodale (personal communication) reports that a higher proportion of Mexican-American males than females enroll in advanced mathematics courses and Gemmill et al. (1982) report a similar sex effect among highly able Mexican-American students, so few Mexican-Americans

of either sex pursue advanced mathematics study in high school and college that sexual differentiation in this subpopulation may be difficult to assess accurately.

In Mathematics Knowledge, the performance of males and females is almost identical in both the poor Hispanic and nonpoor Black subsamples. This suggests that the mathematics training of the sexes in these groups is very similar. However, poor Blacks show as much sexual differentiation in performance, in favor of males, as seen in the two income categories of Whites and the nonpoor Hispanic subsample, amounting to about 20 points. An explanation suggested by Sewell et al. (1982) for their finding that Black inner-city adolescent males obtain higher Peabody Picture Vocabulary Test (PPVT) scores than their female counterparts may apply for the sex difference in mathematics knowledge seen here: the home environment of poor youth may not give females the special incentives for achievement that seem to be characteristic of middle-class Blacks. Moreover, home demands competing and interfering with school achievement among poor Blacks may fall more heavily on girls than on boys.

Verbal Attainment. As seen in the pattern of sex and education interactions, there is very little difference in the performance of males and females in the Word Knowledge test within the various categories of young people. Differences that are seen occur in favor of females in all the groups except poor Blacks and nonpoor Hispanics. In these subsamples, the difference favors males. In the case of nonpoor Hispanics, we speculate that the more home-oriented life style of females, which gives them fewer occasions than males to use English and acquire English vocabulary, contributes to this sex effect. Barral's (1977) investigation of the relative scholastic achievement of foreign-born and native-born Mexican-American students suggests that this sex role proscription may be a factor. Barral reports no significant

difference in the school achievement of foreign-born and native-born Mexican-American children; however, home use of English was a significant determinant of females' school performance but not males'. Although this investigator did not advance an explanation of this finding, it is consistent with the fact that females in this group are expected to stay closer to home and are more protected in contacts with individuals outside the family (Murillo, 1976). That we do not see this sex effect in the poor Hispanic subsample in Word Knowledge probably reflects the greater linguistic isolation of both sexes in this income category as compared to the nonpoor. Kulvesky (1981) reports that the Mexican-American males included in his Texas sample tend to be more Spanish dominant in their oral language with parents, close friends in the neighborhood, and friends in school than is true of females. This suggests that male peers outside the home may also be Spanish dominant. Therefore, though poor Mexican-American males may be encouraged to gain knowledge and experience outside the home (Murillo, 1976), their language experiences may be primarily Spanish. The same may be true for others in the poor Hispanic subsample besides those of Mexican origin.

The sex effect seen for the poor Black group in Word Knowledge is consistent with observations of other investigators. Asbury (1973) reports that among first-graders in rural North Carolina, black males scored significantly higher than females in the Peabody Picture Vocabulary Test (PPVT) and the Metropolitan Readiness Test (reading readiness test). Sewell et al. (1982) indicate that poor Black adolescent males in Philadelphia also score significantly higher in the PPVT than their female counterparts. Sewell et al. speculate that the greater social freedom of males in depressed communities, as compared to that of females in such communities, affords more opportunities to develop cognitive skills, including vocabulary skills, in contexts

that extend beyond home and school. That is, the sex effect seen for poor Blacks may be attributable to some of the same factors that produce somewhat greater Word Knowledge among nonpoor Hispanic males.

Folb's (1980) study of the language and language development of poor Black youth suggests a similar conclusion. Folb reports that there is a considerable pressure on poor Black male adolescents to master the special lexicon of their peer group--a lexicon that evolves very rapidly. There is less of this pressure on girls in these poor communities and thus less incentive for them to broaden their word knowledge.

In Paragraph Comprehension, the performance of poor Black males and females is identical. This test also shows poor Black and nonpoor Hispanics to be similar in the pattern of sex effects. In both these groups, the average performance of males and females is essentially the same, while in other categories of young people, females exceed males in Paragraph Comprehension by about 20 scale points.

In the case of nonpoor Hispanics, families are more likely to possess the economic resources and provide the necessary encouragement to insure the training of males and females alike in this basic skill. The parental encouragement that appears so critical for the scholastic achievement of Hispanic young people (Brown et al., 1980; Gemmill et al., 1982) may not be as strong among poor Hispanics. Thus, the development of reading ability in low income Hispanics may be more determined by the factors that contribute to the sexual differentiation seen in this skill among Whites.

The similarity of the scores of poor Black males and females in Paragraph Comprehension is another expression of the disadvantaged position of males among the poor Blacks already seen for Word Knowledge. In the case of Paragraph Comprehension it nullifies the expected superiority of females and leaves the sexes equal in performance.

### 7.3 Sources of Sex Effects

There is a wealth of literature on the possible origins of sex differences in test performance. We review the main results in this section. (An hypothesized biological variable, however, sex difference in spatial ability, is examined separately in Section 7.4.)

Sex Differences among Blacks. Our results go considerably beyond previous findings indicating that Black females show a higher level of school achievement and educational and occupational attainment than Black males (Jackson, 1973; Edwards, 1979). It is true, in the current sample, that substantially larger proportions of the Black females have completed some college (See Table 2.5-8) than is true for Black males. But the test data suggest that the educational attainment differences seen between Black females and males is not the result of the females having uniformly superior skills relative to their male counterparts. Black males, regardless of income level, exceed the performance of Black females in the five specialized knowledge tests (i.e., General Science, Arithmetic Reasoning, Auto and Shop Information, Mechanical Comprehension, and Electronics Information), although the amount of difference in favor of males in the Black subsamples is generally less than that seen for males in other sociocultural-income groups. Poor Black males exceed the performance of their female counterparts in Word Knowledge, equal their performance in Paragraph Comprehension, and outperform the females in their group in both Mathematics Knowledge and Arithmetic Reasoning. Among nonpoor Blacks, females exceed the performance of their male counterparts in Word Knowledge and Paragraph Comprehension and equal males' performance in Mathematics Knowledge though Black males show an advantage in Arithmetic Reasoning. These data support Ogbu's (1978) hypothesis that Black females show higher scholastic achievement and educational attainment than their male

counterparts because they are more motivated to achieve scholastically. In other words, Black females' attainment may be more the result of greater effort and persistence in educational contexts than of higher level skill development as measured by the tests included in the ASVAB. Astin (1982) reports from his analysis of factors affecting minority higher educational progress that "being a woman is positively related to persistence among Blacks, but negatively related among Chicanos and Whites" (pp. 93-94).

Non-poor Black females' educational motivation appears to be conditioned by the greater family support they receive for educational attainment as compared to males in this group. Blau (1981) reports from interviews with 579 Black and 523 White mothers of Chicago fifth- and sixth-grade children in 1968 that Black mothers generally expressed higher ambitions for their daughters than sons, while White mothers espoused higher ambitions for sons. Both Edwards (1979) and Ogbu (1978) posit that Black families have, historically, invested more of their resources in the education of daughters than sons because the employment opportunities have been more favorable for Black females than Black males. As indicated by Edwards (1979), however, the gap in the educational attainment of Black males and females appears to be closing. Analyzing sex and cohort changes for Black college graduates between 1940 and 1970, Edwards attributes the increase in Black male college graduates to economic advancements for Blacks in recent years and changes in the opportunity structure which portend better occupational prospects for Black males with college degrees.

Similarly, Thomas's (1980) analysis of data from the National Longitudinal Study (NLS) of the High School Class of 1972 revealed that the direct effects of family status on college entry were greater for male and female Blacks than for Whites and greatest of all for Black males. This suggests

that the Black family's investment in the education of males remains more strongly determined by family resources than is true for the investment in females.

Sex Differences in Fluent Production. Female superiority in verbal fluency, fine motor movements and repetitive tasks is universally acknowledged. It appears clearly in the Numerical Operations and Coding Speed tests of the ASVAB, though relatively reduced in the poor Black group. Despite the long time that this phenomenon has been known, there has been relatively little research into its causes or consequences. Broverman and Klaiber (1969) are the only psychologists to have given it prominence. They note that, among persons of the same general verbal ability, spatial ability scores and fluent production scores are negatively related. Petersen (1976) confirmed this finding. Although verbal, spatial, and fluent production tests load on different factors (Bock, 1973), the correlations among the factors are such that negative partial correlations of spatial and fluency measures, holding fixed verbal ability, can occur. Broverman and Klaiber advance a hormonal theory, with mediation by brain monoamine oxidase (MAO), to explain the sex differences and individual variation within sexes. More sexually differentiated individuals of either sex (i.e., more somatically masculinized males and more somatically feminized females) are supposed to have lower MAO levels resulting in higher brain norepinephrine, better fluent production, but reduced analytical ability.

An alternative explanation for females' advantage in fluent production tests, particularly as reflected in the Numerical Operations and Coding Speed tests of the ASVAB, is that females may be more responsive to the demand for speed in tasks such as these. There is some evidence to suggest that females, at least in the early years, are more likely than males to comply with the directions of authority figures such as parents, teachers, experimenters in

laboratory situations, and examiners in individual test administrations (Serbin et al., 1973; Minton, 1971; Hertizig et al., 1968). It is also hypothesized (Anastasi, 1982) that one reason investigators consistently find higher correlations for females than males between aptitude test scores and academic grades (particularly among college students) is that females may be more inclined to conform to the values and achievement standards of class contexts such that their achievement depends more on their abilities. Anastasi suggests that males may be more discriminating in their achievement behavior, choosing to persist in tasks and subject matter that arouse their personal interest. While these observations are speculative, socially-conditioned sex differences in young people's willingness to conform to task requirements and/or motivation to perform certain types of tasks cannot be eliminated as potential sources of variance in the fluency test scores.

Sex Differences in Mathematics Achievement. The differences in males' and females' mathematics achievement seen in the Profile data are similar to those found in previous investigations (Fennema, 1980; Armstrong, 1980, 1981; Centra, 1974). The literature indicates that sex-related differences in mathematics achievement are not consistently found until early adolescence (Fennema, 1974; Sherman, 1977; Maccoby and Jacklin, 1974). Generally, adolescent females outperform males in formal mathematical operations, while males show higher performance in less formal and more analytical reasoning tasks (Fennema, 1974). Armstrong (1980) reports from an analysis of the mathematics achievement of a national sample of 13-year-olds and high school seniors that the female 13-year-olds achieved higher scores than their male counterparts in arithmetic computations; however, no sex-related differences were seen in problem-solving ability in this age group. Among high school seniors, no sex differences were seen in computation or algebra achievement, but a large and



statistically significant difference, favoring males, was observed between the sexes in problem-solving ability. Armstrong concludes from this analysis that males and females enter high school with essentially the same level of skill and understanding in mathematics, but during the high school years males surpass females in their ability to solve one- and two-step "word problems."

Why males show higher performance in mathematical reasoning than females during the high school years is a question that has stimulated a considerable amount of research in recent years. Mathematics competence greatly enhances the educational and career options available to young people. Therefore, the fact that females do not develop mathematics mastery during the secondary school years in the same proportions as their male peers has been identified as a critical factor in the persistent underrepresentation of women in science and mathematics-related fields, as well as the technical professions (Sells, 1980; Sherman, 1982; Fennema, 1982; Meece, 1980).

An important factor bearing on this problem is the relative proportions of males and females who enroll in advanced mathematics courses (i.e., college track mathematics) in high school. Studies reviewed by Fennema (1982) show considerably fewer females than males enrolled in advanced mathematics courses in secondary school. On these grounds, sex differences in mathematics achievement were thought to result primarily from the fact that females less often elected to study mathematics than males.

The data presented in this report show higher average scores for males not only in mathematics and mechanical comprehension, but also in electronics information and science knowledge--areas that are highly influenced by education. Although sex differences in these areas of specialized knowledge have not been the specific focus of psychological research, it is generally thought that the factors contributing to male advantages in mathematics also affect

these related areas (Fennema, 1981). Thus, the research findings related to sex differences in mathematics may also suggest explanations for those observed in the other specialized knowledge tests of the ASVAB.

Numerous variables have been implicated as factors that influence female attitudes and achievement in mathematics, including sex-typing of mathematics as a male domain (Fox et al., 1979), sex differences in perceived causes of success and failure (Fennema, 1981; Wolleat et al., 1980), differential parental expectations of sons' and daughters' success in mathematics (Levine, 1976; Casserly, 1980), sex differences in perceived usefulness of mathematics (Haven, 1971, Sherman and Fennema, 1977), sex differences in self-concept of mathematics ability (Brush, 1980; Sherman, 1980), differential treatment of males and females by teachers and counselors (Fennema, 1980; Casserly, 1980), sex differences in self-confidence in relation to specific tasks (Maccoby and Jacklin, 1974), differential early socialization of males and females for achievement orientations (Rubovits, 1975), sex role orientation (Bem and Lenny, 1977; Nash, 1979), and actual fear of success among females (He ner, 1972). All of these factors tend to work in the direction of reduced female participation in advanced mathematics courses and, thus, influence achievement.

Findings from two recent large-scale investigations question the assumption of differential course participation. The Women in Mathematics Project, a survey conducted by the Education Commission of the States (ECS) in the Fall of 1978 with over 1,700 twelfth grade males and females nationwide, and the second Mathematics Assessment, conducted during the 1977-78 school year by the National Assessment of Educational Progress (NAEP) with a national probability sample of 75,000 9-, 13-, and 17-year-olds, show near equality in the mathematics course participation of the sexes (Armstrong, 1981). The

Women in Mathematics Project data indicate no significant difference in the proportions of males and females who were taking or had completed courses in calculus, precalculus, trigonometry, geometry and computer programming. As a matter of fact, of the twelve courses covered in the survey, only two showed a significant sex difference in favor of males (probability/statistics and algebra II), and one showed a significant difference in favor of females (accounting/business mathematics). The NAEP study also revealed few sex differences in the mathematics course participation of 17-year-old students. Of the nine courses surveyed in this study, only two showed a significant difference in the proportions of males and females who had studied the course--precalculus/calculus and trigonometry. The NAEP survey also collected data from the 13-year-olds regarding the number of years of high school mathematics they planned to take. Significantly more males than females planned to take only one year of high school mathematics, but no sex differences were seen for students planning two, three, or four years of mathematics training.

Nevertheless, sex differences in mathematics achievement favoring males were seen in both studies among 17-year-olds and 12th graders. The males at every level of course participation exceeded the performance of females in solving typical one- and two-step word problems. Armstrong (1981) concludes from this analysis that "It is clear that achievement differences are not solely a function of differences in participation" (p. 371). More recently, however, Pallas and Alexander (1983) examined coursework effects in a prospective study of sex differences in mathematics achievement. Using the data from the Educational Testing Service Academic Prediction and Growth Study (Hilton, 1979), they examined the scores of 6,119 12th grade public school pupils who took the mathematics subtest of the PSAT (SAT-M) early in 1968. Because these pupils were part of a longitudinal study, their quantitative

Cooperative School and College Ability Test (SCAT-Q) scores in 9th grade, as well as records of their intervening course-work, were also available. Pallas and Alexander found that, although these pupils entered high school showing practically no sex difference in mathematics attainment, by 12th grade the difference was substantial. Partial regression analyses showed that these differences in math achievement are strongly affected by course enrollment, especially enrollment in advanced mathematics courses, and that this effect could not be eliminated by controlling on 9th grade SCAT-Q scores. The analysis indicated that, while not all of the sex differences in 12th grade SAT-M scores could be attributed to sex differences in mathematics course enrollments, the portion that could be so attributed was large enough to leave little doubt that the differential coursework hypothesis has a basis in fact. The authors suggest that other hypotheses would have to be considered to explain the residual sex difference not attributable to course enrollments. Since the recent studies suggest that mathematics course participation is not the only factor in observed sex differences in achievement, there is an obvious need to consider personal characteristics of males and females as learners of mathematics as potential sources of variance in mathematics achievement of the sexes.

Since the perceived usefulness of mathematics does affect mathematics participation and achievement, the near equality in the high school mathematics course participation (Brush, 1980; Sherman and Fennema, 1977) of males and females noted in the Women in Mathematics and NAEP surveys (Armstrong, 1981) suggests that the sex-related difference in perceived usefulness of mathematics may be decreasing. In other words, the task value of mathematics for females may be increasing, at least as indexed by mathematics course participation. It is not clear, however, whether this trend indicates a

substantial increase in the proportion of females who are planning careers in science and mathematics-related fields. Matthews (1980) reports from a survey of students enrolled in advanced mathematics courses that twice as many White males (86 percent) as White females (43 percent) indicated as an important reason for taking mathematics that "it is important to my career." In fact, a higher proportion of the Black (56 percent) and Asian (80 percent) females surveyed in Matthews' study reported career plans as an important factor in their decision to enroll in advanced mathematics courses than was true for White females. Matthews' data suggest that the increase in females' persistence in mathematics training in high school may be determined more by their plans for higher education than plans to pursue careers in science and mathematics.

A review of the literature in this area, guided by the Meece et al. (1982) model of individual academic choice, suggests that success expectancy in mathematics may be a particularly significant factor in the disparities seen between males and females in mathematics course enrollment and achievement.

Meece et al. (1982) assume that the underrepresentation of women in certain courses of study, specifically advanced mathematics, occurs because low success expectancy leads them to choose their training in other areas. However, in view of the considerable differences between the sexes in the affective and attitudinal variables related to mathematics participation that are apparent by junior high school, one may question whether the difference in the programs of study pursued by males and females is, in reality, a matter of choice, which implies selecting from a variety of alternatives after some deliberation. Rather, the presence of sex-role proscriptions, and the socialization process by which they are inculcated, may act to restrict the

educational and occupational options available to women to such a degree that these decisions are more a matter of acquiescence than of choice.

It has been theorized that occupational role choices, and the related educational choices, are so closely tied to fundamental definitions of sex identity that nontraditional choices in these areas may cause considerable identity conflict (Rosenberg and Rosenberg, 1978; Nash, 1979; Fox, Tobin and Brody, 1979). Nash (1979) suggests that the need for cognitive consistency in sex role identity is a mediator of intellectual achievement, particularly during adolescence when belief in the appropriateness of one's sex role behavior becomes critical for the consolidation of ego identity. This view is supported by interviews in which Sherman (1982) investigated attitudes toward sex role and mathematics among girls who elected to take varying amounts of theoretical math in high school, but were comparable in general ability. The results of these interviews indicate that girls who enrolled in four years of mathematics, more than any other group, showed evidence of sex-role conflict. However, the girls who took four years of mathematics were more ambitious than those who had taken only three years of mathematics, which raised the following question for Sherman. "Did the girls who had taken less math show less ambition because they had resolved aspects of sex-role conflict in favor of renouncing ambition?" (Sherman, 1982, p. 440). This question could not be answered empirically from the interview data available for the girls, in this study, but Sherman speculates that this factor did play a role in the girls' decision regarding mathematics course participation.

Sherman's (1982) findings of greater sex-role conflict among girls who persist in the study of theoretical mathematics is consistent with the work of other investigators who have attempted to identify the salient aspects of adolescence which contribute to the emergence of sex-related differences in

cognitive achievement, particularly mathematics achievement. This research indicates that the formation of heterosexual relationships during this time may be particularly important (Fennema and Sherman, 1977; Nash, 1979). Levine (1976) reports that girls generally believe that boys do not like smart girls, particularly those who excel in mathematics. Sherman's (1982) study confirms this finding and its possible relationship to females' mathematics participation decision and achievement. Fear of peer rejection, in general, has also been identified as an important factor in mathematically precocious girls' decision to not pursue advanced study in mathematics (Fox, 1977). This literature generally questions the extent to which females actually have a choice in their pursuit of various educational programs and careers given their perception of the psychosocial cost of deviation from the norms set by the culture.

There is some evidence to suggest that teachers display lower expectations for females in mathematics achievement than they do for males and that school counselors actually discourage girls who are interested in mathematics and science careers by advising them against advanced courses in these areas (Haven, 1971; Casserly, 1980; Luchins, 1976). That this may be the case is suggested by Gregory's (1977) finding that teachers refer boys for remedial instruction in mathematics significantly more often than they refer girls. Parsons et al. (1982) report a significant relationship between teachers' expectations for a student and the student's self-concept of mathematics ability, regardless of past mathematics achievement (i.e., grades). However, the extent to which teachers' expectations for young people as learners of mathematics are operationalized in different interaction patterns in the classroom was not clear from this investigation.

To summarize to this point, the research literature generally shows that the process of sex role expectations condition behavior in both social and intellectual areas in keeping with sex-appropriateness. The extent to which a particular task or subject matter is viewed as masculine or feminine reinforces sex differences in performance and development of competence, which appear to be mediated by differential attainment values, confidence and persistence.

#### 7.4 Sex Differences in Spatial Ability

Some investigators have suggested that sex differences in spatial visualizing ability--the ability to manipulate images without the aid of verbal mediation (Peterson, 1979)--play an important role in observed differences in mathematics reasoning. Although researchers have yet to develop consensus on the nature and significance of the relationship (Fennema, 1980; Sherman, 1980; Benbow and Stanley, 1980), the literature suggests that the sex differences seen in the Profile data for Arithmetic Reasoning and Mechanical Comprehension may reflect the influence of sexual differentiation in spatial visualizing ability. For example, when presented with a configuration of interlocking gears and rods, prediction of the movement of contiguous parts would be facilitated by the ability to manipulate the configuration visually. Without this ability, a lengthy concatenation of verbal propositions is required to reach conclusions that are obtained immediately by visual imagery. The same may be true for the solution of one- and two-step word problems, such as appear in the ASVAB Arithmetic Reasoning test. In mathematics problems of this sort, quantities are presented in an unstructured form and in an arbitrary order that must be brought into a configuration that has a mathematical representation. As we discuss below, spatial visualization of the configuration may facilitate this type of thinking.



Although the Arithmetic Reasoning and Mathematics Knowledge tests loaded on the same factor in the factor analysis reported in Chapter 3, they do not show the same sex differences. As stated above, sex differences in Arithmetic Reasoning appear at all educational levels, while those for Mathematical Knowledge are almost nonexistent except for a minimal difference at the highest educational level. The explanation of this seeming contradiction is that the factor analyses are describing the individual-difference variation within same sex groups, whereas the analysis of group means reveals the often small but systematic effects that make one group of subjects different from another on average. The factor analysis shows that persons who are proficient in Arithmetic Reasoning also tend to be proficient in Mathematics Knowledge, but it takes multivariate analysis of variance to detect the systematic tendency of male subjects to exceed female subjects in Arithmetic Reasoning but not in Mathematics Knowledge.

Other investigators have called attention to the distinction between tests that require the solution of quantitative problems versus those that require recall of mathematical information or the execution of formal arithmetic or algebraic rules. Beginning in adolescence and continuing to adulthood, sex differences in favor of males appear in the problem-solving tasks but not in the information or operations tasks. Werdellin (1961), for example, working with high school pupils, found better female performance on geometry information items and no sex differences on items that could be solved by well-known formulas or in a step-by-step fashion. But on items that required the solution of problems involving geometrical principles or reorganization of relationships, that is, analysis, he found clear male superiority. A study by Armstrong (1980) yielded similar results for computation and algebraic operations vs. algebraic problem-solving. The former did not show sex differences

among high school seniors, whereas the latter showed large significant differences favoring males. Eighth grade pupils in this same study, in contrast, did not show sex differences either in arithmetic computation or problem-solving. This study corroborates the typical finding that sex differences in quantitative reasoning first became marked when adolescent changes become apparent in boys at about age 14. In studies where consistent sex differences in mathematics attainment are not found, such as those of Fennema and Sherman (1977), achievement tests have been used that emphasize mathematical information and content rather than quantitative problem-solving.

As early as 1940, it was suggested by Gastrin (1940) that sex differences in geometrical ability could be attributed to the male superiority in the factor of spatial visualizing ability that had been found in factor analysis of mental tasks in the 1920s and 1930s. McFarlane (1925) found in a set of "practical ability" tests a group factor distinct from the general verbal ability factor and having something to do with judging the relationships between physical objects. She noted that boys surpassed girls on these tests. By 1928, Truman Kelley had devised objective test items measuring spatial ability and had identified by principal component analysis a spatial factor requiring the mental manipulation of shapes.

Spatial tests were prominent in the batteries developed for pilot selection during World War II (Guilford and Zimmerman, 1947). Excellent predictive validities for this type of test have been found in validation studies. Not surprisingly, spatial tests predict skill in engineering drawing far better than do tests of verbal ability (Holliday, 1940, 1943; Slater, 1941). In other practical areas such as shop training and auto mechanics training, spatial tests have good predictive validity (Martin, 1951). The 1974 manual for the Differential Aptitude Tests (Bennett, Seashore, & Wesman,

1974) reported predictive validities of its spatial relations test of  $r = .42$  for drafting,  $r = .47$  for shop mechanics and  $r = .69$  for watch repair. Recently, Gibbons (1981) obtained a correlation of .7 between ratings of residents' manual surgical skill and a measure closely related to spatial tests (Bender Gestalt). Predictive validities of spatial tests have recently been reviewed by McGee (1979).

A striking property of spatial tests is the large sex differences they exhibit when administered to both adolescent and mature subjects. In contrast with tests of verbal ability, which show either no sex difference or a slight female superiority, the sex difference in spatial test performance favoring males is one of the most consistent and well-established findings in cognitive psychology. Even simple two-dimensional versions of the tests, such as the Thurstone and Thurstone (1947) item, shown in Figure 7.4-1, are easier for male subjects, and still greater differences in the same direction are found for three-dimensional versions such as the identical blocks test, an item from which is shown in Figure 7.4-2 (Stafford, 1961). The distinctive feature of spatial visualization items (which show strong sex differences and load on the spatial visualization factor) is that the problem they pose cannot be readily expressed in verbal form. It can be solved most quickly by directly imagining

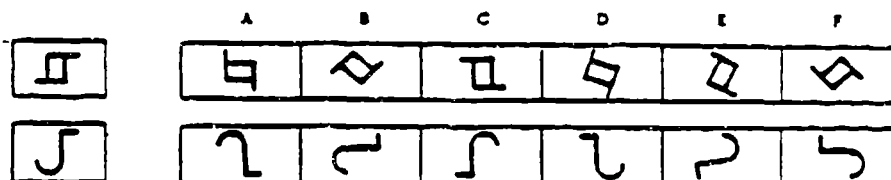


Figure 7.4-1 Two items from the Thurstone and Thurstone Primary Mental Abilities Test (1947)

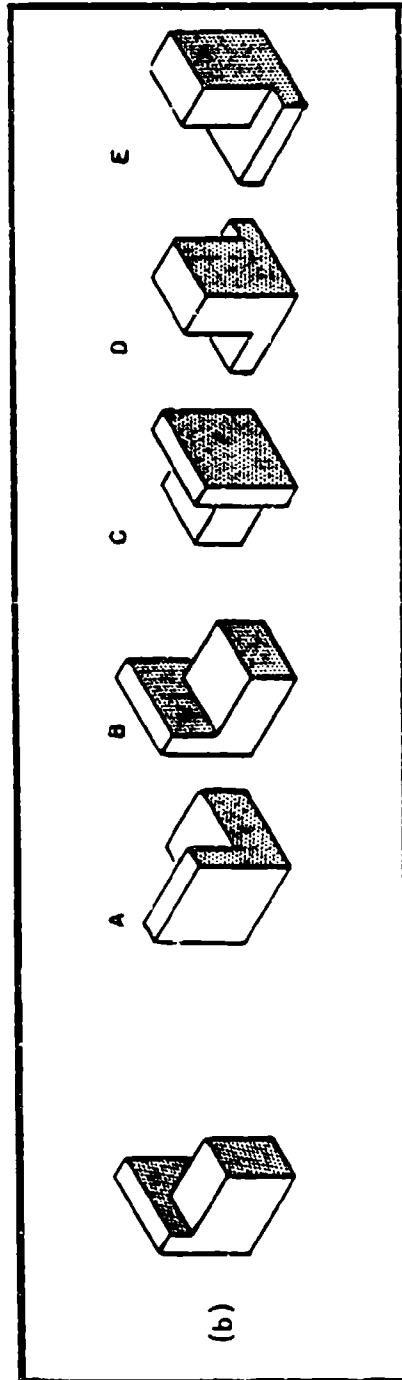


Figure 7.4-2 Identical Block Test (Stafford, 1961)

the appearance of the object after it is turned. Studies by Shepard and Metzler (1971) have found that the time required to decide whether rotated objects, such as shown in Figure 7.4-3, are identical (and not mirror images) is proportional to the number of degrees that the objects are rotated from one another. This would seem to imply that the subjects are performing an analogue mental rotation in real time. (Subjects who are able to solve these types of items quickly give subjective reports of processing the images in this way.) These results suggest that significant cognitive problems can be solved without verbal mediation and that productive thought can consist of transformation of images rather than recall and relating of propositions. This is not to say that these two kinds of thought cannot take place simultaneously; the introspecting subject can describe an image in terms of features and orientations that have names and descriptors in the language. To the extent that an item presented in the form of a figure can be solved by naming distinctive features and noting their positions, rather than manipulating an image of the configuration as a whole, the item becomes verbal rather than spatial and does not show appreciable sex differences.

That the human mind is capable of carrying out verbal and spatial thought separately has been substantiated by the remarkable finding that these functions are served by distinct parts of the brain. Studies of brain damaged aphasic subjects (Geshwind, 1974) and of so-called "split-brain" subjects who

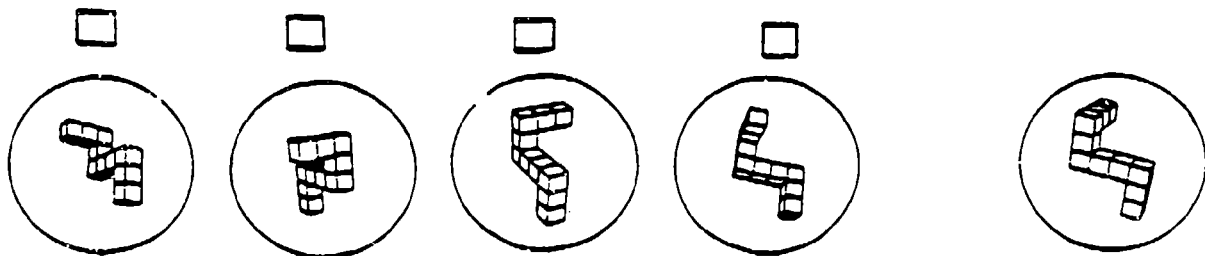


Figure 7.4-3 An item from the Vandenberg-Sheppard Mental Rotation Test

have undergone sectioning of the corpus callosum and anterior commissures to control epileptic attacks (Sperry, 1970; Gazzaniga, 1970; and Nebes and Sperry, 1971) confirm that in the vast majority of people the capacity to name objects and utter sentences is the exclusive province of the frontal and temporal lobes of the left hemisphere. The right hemisphere can recognize objects and even transform images but is mute. It cannot name the object but can demonstrate by gesturing or pointing with the left hand that it knows the appropriate use of the object. When a familiar object is placed in the finger tips of the left hand of a blindfolded split-brain subject, the brain receives information only from the projection of sensory nerves in the finger tips onto the right hemisphere. The split-brain subject cannot then name the object but can perform gestures appropriate to its use.

Other evidence supporting this interpretation is seen in cases of brain damage. Lesions in the left hemisphere frequently interfere with speech, but the capacity for spatial problem-solving remains intact. Conversely, lesions in the right hemisphere depress spatial ability in most patients without affecting speech (Geshwind, 1980).

Hemispheric specialization of function also explains the finding of Kimura (1967) and others that in normal subjects the right ear, the innervation of which projects most directly on the left hemisphere, has a slight superiority for discriminating speech sounds while the left ear has an advantage for non-speech sounds. Similarly, electroencephalographic studies (Ornstein, 1973) indicate that increased brain activity in the right hemisphere occurs during spatial tests, and in the left during verbal tests.

Rock (1973) believes that this anatomical separation of function is almost certainly the source of the verbal and spatial factors in mental test scores. Individual differences in the vascular and neural architecture of

different parts of the brain are commonplace anatomical observations. These differences would be expected to affect differentially the two sides of the brain and produce relatively independent variation in the speed and effectiveness of verbal and spatial thought. Whether this functional specialization of the hemispheres could also account for sex differences in cognition is currently being debated. From studies of brain-damaged patients, McGlone (1980) has presented evidence that the hemispheric specialization of verbal and spatial information processing is less complete in women than in men. In the patients she studied, right hemisphere lesions sometimes affected language in women but rarely in men. Although her results have yet to be confirmed by other investigators, such differences in hemisphere specialization could perhaps explain the sex difference in spatial ability on the grounds that the presence of linguistic functions in the right hemisphere interferes with spatial thought. This interpretation will perhaps be clarified as recently developed techniques of monitoring activity at specific sites in the brain, through measurement of local blood supply or by improved electroencephalography, make it possible to distinguish between verbal and spatial thought in normal subjects.

Hormonal Influences. Sex difference in spatial visualizing ability may have a developmental physiological component. Waber (1977) found some evidence that later maturing boys and girls in the 13 to 15 year age range had higher spatial scores than earlier maturers. Since onset of puberty is triggered by increasing estrogen levels in girls and testosterone levels in boys, the implication is that persons who are hormonally more strongly sexually differentiated at an earlier age are not as adept in spatial tasks as less differentiated persons. This means that the better spatial visualizing females would tend to be less feminized somatically and the comparable males,

less masculinized. This is what Petersen (1976) observed in her study of spatial visualization vs. fluent production in young people from the Fels longitudinal study. It suggests that there is an intermediate range of androgen-estrogen balance that is favorable to the development of spatial visualizing ability. The fact that there are more good spatial visualizers among males would then be explained by more males falling in this critical range. Since there is evidence that the timing of puberty is heritable (Bock and Thissen, 1978), this interpretation of variation in spatial ability would not be inconsistent with the genetic research.

Other evidence of a biological basis for individual differences in spatial visualizing ability comes from family studies, virtually all of which show that proficiency on spatial tests is substantially inherited. Family and twin studies have consistently shown heritabilities for spatial performance comparable to that of the more widely investigated verbal abilities and general intelligence. Vandenberg (1967) reviews many of these results. Moreover, factor analytic studies of the heritable variation in batteries of cognitive tests show that the heritable spatial component is distinct from the heritable verbal component (Bock, 1973; Vandenberg, 1975). Although all of these studies have been based on parent and offspring or dizygotic or monozygotic twins living together, the rather esoteric nature of the spatial trait makes it unlikely that the transmission from parent to offspring could be due to some form of imitation independent of verbal learning. It seems similarly unlikely that the environments of monozygotic twins living together could be more similar with respect to sources of the spatial trait independent of the possible greater similarity of the verbal environment of monozygotic twins. [Monozygotic twins are known to spend more time together and to interact verbally more than dizygotic twins: for a recent review, see Segal (1981).]



Further evidence of genetic influence on the spatial trait comes from observations of bimodalities in the spatial test score distributions. Such a bimodality was first observed in the scores of a spatial test administered to a large sample of young adult male subjects as part of a factorial study of mental ability (Thurstone and Thurstone, 1941). None of the other tests in this study showed any sign of bimodal score distribution--a rare event for any sort of cognitive measure. A similar result was obtained by Bock and Kolokowski (1973) by resolution of score distributions of the Guilford-Zimmerman (1953) spatial visualization test into Gaussian components. Subsequent studies have confirmed the bimodal nature of the spatial score distributions (Loehlin et al., 1978; Yen, 1975 Ashton et al., 1979). These results suggested that the subjects could be classified as "visualizers" or "non-visualizers" corresponding to two similar normal distributions separated by about one standard deviation. In male subjects, approximately equal numbers were visualizers and non-visualizers, but in female subjects, only about 25 percent were visualizers. This finding agrees with the observations of many workers that in adult subjects about 25 percent of females score above the male median in tests of spatial ability (O'Connor, 1943; Bouchard and McGee, 1977). The fact that average spatial scores are typically lower in women can be attributed to the smaller number of women who are "visualizers"; women with the visualizing trait score just as high on spatial tests as do the male "visualizers."

On the basis of these observed proportions of visualizers among males and females, O'Connor (1943) proposed that spatial visualizing ability is due to an X-linked recessive gene with a frequency of about  $1/2$  in the population. Such a trait would be expressed by half the males (whose cells carry only one chromosome) and  $1/2 \times 1/2 = 1/4$  of females, who must inherit the recessive visualizing gene from both parents in order to express the trait.

Another consequence of X-linkage of a recessive trait is a characteristic correlation among family members classified by sex. If a test score is influenced by an X-linked recessive gene, then among parent and offspring the correlation of fathers and sons should be zero, that of mother and son and father and daughter should be greatest, and that of mother and daughter intermediate. Between offspring, sisters should show the highest correlations, brothers and sisters the lowest, and brothers and brothers intermediate. The numerous family studies of spatial visualizing abilities since Stafford's (1961) initial effort have been reviewed by McGee (1979). The results to date are highly variable. Some studies show a pattern of correlation suggestive of sex-linkage (Stafford, 1961; Hartlage, 1970; Bock and Kolokowski, 1973; Yen, 1975; Walker, Krasnoff and Peoco, 1981). Other studies show the pattern suggestive of sex-linkage for one of the tests used but no such pattern for another (Park et al., 1978; Loehlin, 1978). Still others show a distinctly contrary pattern (DeFries et al., 1976; Bouchard and McGee, 1976). It appears that, unlike the more robust distributional results, the correlational studies are so sensitive to the spatial tests employed or other differences in experimental procedure that a clear result with respect to X-linkage cannot be obtained by this method. A much more decisive approach would be to look for linkage between the putative major gene for spatial ability and a marker locus on the X chromosome (Bock and Perline, 1977). This has in fact been attempted with positive results for sex linkage by Goodenough et al. (1977), but the sample size was too small to give any real confidence in the results.

Another genetic theory that would explain differences in spatial ability without resorting to X-linkage is x-influence; that is, the expression of a dominant, somatic major gene for the ability is influenced by differences in sex hormone levels among the subjects. If there is some threshold of

androgen/estrogen balance beyond which the trait is not expressed, it may be that virtually all male visualizers are above this threshold but half of the female visualizers are not. This would explain the proportion of visualizers and nonvisualizers seen in the Gaussian resolution of the spatial scores.

Evidence supporting this X-influence theory is seen in subjects suffering from Turner's syndrome. These women are completely agonadal and exhibit, in the presence of virtually normal verbal intelligence, a specific and striking deficit in spatial visualizing ability (Garron, 1970). Similarly, male subjects who show gonadal hypoplasia at adolescence perform at female levels on spatial tests (Hier and Crowley, 1980, 1982). Finally, there is an observation suggesting that adolescent girls who are visualizers exhibit some tendency to more masculine physique (Petersen, 1981).

If spatial ability indeed has some partly heritable physiological basis, it would mean that the biological specialization of the two sexes with respect to reproductive function also extends to cognitive functioning. Because of the considerable variation within the sexes, however, little can be said about the spatial ability of a particular man or woman without specifically testing for the trait. Many women are visualizers and can be identified by the use of the tests. It is only in the distribution of scores in the population that a sex difference in spatial visualizing ability actually exists.

Spatial Ability and Quantitative Reasoning. That the sex difference in performance on quantitative reasoning is due to differences in spatial ability problems has been argued by Smith (1964) and Sherman (1967); they did not, however, have the evidence to establish the point conclusively. But in 1975, Hyde, Geiringer and Yen investigated in a sample of 45 male and 35 female undergraduates the relationship between scores on the Identical Blocks test

(see Figure 7.4-2), the Rod and Frame test, the Group Embedded Figures test (Witkin et al., 1962), a mental arithmetic test (Stafford, 1965), and the WAIS Vocabulary subscale. Using the Identical Block score as a covariate, Hyde et al. (1975) were able by the analysis of covariance technique to eliminate the significant difference favoring males both on the mental arithmetic and the Rod and Frame test. This suggests that these differences also arose from the spatial trait. Although a small difference favoring males remained after the covariate adjustment, this is to be expected if the covariant is fallible, as all test scores necessarily are.

In a similar study, Burnett et al. (1979) examined the relationship between the Scholastic Aptitude Test (SAT) Quantitative Scale and spatial ability as measured by the Identical Blocks test and the Guilford-Zimmerman Spatial Visualization Test. In a group of 183 male and 81 female college students selected for their mathematical ability, they found that males scored significantly higher on both the spatial test and the SAT Quantitative, while females outperformed males on the SAT verbal. Correlations between the spatial tests and the SAT Quantitative in the two groups ( $r = .40, .57$ ) were almost as high as between the two spatial tests ( $r = .50, .57$ ), whereas the correlation between the spatial test and the SAT verbal were comparatively low ( $r = .07, .28$ ). As in the Hyde et al. study, analysis of covariance with the two spatial scores as covariants reduced the sex difference in the SAT Quantitative scores to insignificance.

These studies support the conclusion that the sex differences observed with quantitative reasoning tests arise from the corresponding difference in spatial ability. Only the Fennema and Sherman (1977) study gave mixed results in this connection, but they were using a general test of mathematics achievement (The Scardell Test of Academic Progress, 1972) including many aspects of

mathematics knowledge where sex differences are not expected. Indeed, Benbow and Stanley (1980) argue that the Scardell test is not a good measure of mathematical aptitude. On these grounds, the Fennema and Sheiman study does not seem a serious challenge to the spatial interpretation of sex differences in quantitative reasoning.

If there is any reason to question the theory, it might be one's difficulty in understanding how spatial items (see Figure 7.4-1, 7.4-2, and 7.4-3) and arithmetic reasoning items (Figure 3.4-1) draw on any common cognitive function. Arithmetic word problems are, as their name implies, almost always presented in words and not in images. Like many of the items on the ASVAB Arithmetic Reasoning test, they may involve quantities of money, which do not normally have a physical representation. All that can be said is that these items present in an unstructured form and in arbitrary order a number of quantities that must be brought into a configuration that is in one-to-one correspondence with some mathematical formula or algorithm. As Bronowski (1947) has observed, mathematics is a language that communicates those aspects of the world that can be described only in configurations that are not readily amenable to the chains of propositions that make up verbal language. Although some of the items in the ASVAB reasoning test seem more "configural" than others, (for example the item showing the largest sex difference in favor of males is a ratio problem involving the height of two sunlit objects casting shadows), even the unit price and extension problems show sex differences favoring males. From thresholds for these items reported by Bock and Mislevy (1981), one can pick out the items where sex differences are greatest. These differences prove not to be related to the content of the items but to difficulty. Items in the latter part of the test involving more subtle application of ratio rules, fractions, and interest calculations show

greater sex differences than the simpler items earlier in the test. Either the female subjects are unable to solve these problems or they are too slow in reaching them or give up before completing the test. Sympson and Weiss (1981) report failure to complete all items of the ASVAB Arithmetic Reasoning Test. But perhaps the real explanation is that arithmetic word problems require active interplay of word and image that puts non-visualizers, in which women are in the majority, at a disadvantage.

## 7.5 Summary and Conclusions

In summary, the analysis of interaction effects between sex and education on the one hand, and sex and sociocultural group on the other, reveals significant sex differences in the vocational preparation of American young people. Males generally exceed females in General Science, Arithmetic Reasoning, Mechanical Comprehension, Electrical Information, and Auto and Shop Information. Females exceed males in Numerical Operations and Coding Speed. The data indicate that, on the average, young American males enjoy a competitive advantage over females for occupations and higher education placements in technical and scientific fields--areas in which females are at present substantially underrepresented--because males have developed higher levels of skills required in these areas by the time they leave high school. It is important to note that the average difference in favor of males in these areas varies considerably with educational level, sociocultural group and income level. These data clearly show that sociocultural group differences in the development of skills required for technical and scientific careers are generally much greater than the sex differences observed within the various categories of young people.

The data generally confirm that sex differences seen among White middle-class youth in mathematical and visual-spatial abilities appear equally

in other social groups, though among Blacks in both poor and nonpoor income classifications and among poor Hispanics the sex differences are relatively smaller but substantial in absolute terms. Sex differences in knowledge of mathematics rules and ability to apply them are much smaller in each of the groups than those seen for Arithmetic Reasoning. In fact, the performance of males and females in the poor Hispanic and nonpoor Black subsamples in Mathematics Knowledge is essentially the same. As is the case for the specialized knowledge tests, the racial differences in mathematics achievement are considerably greater than the sex differences seen within social groups.

The relative advantage of females in verbal ability varies with sociocultural group membership, educational level, and income classification. For the total sample, the female advantage in verbal attainment is seen for respondents with 12 and less years of education. Among respondents with some college, the verbal attainment of males and females is essentially the same.

An interesting finding in the analysis of the sociocultural group by sex interaction is evidence that among poor Blacks, females tend in some respects to be more disadvantaged than males, whereas among nonpoor Blacks, females are relatively more advantaged. We interpret this to mean that stress upon the families of poor Blacks is more descriptive of the school achievement of girls than of boys. Conversely, among Hispanics, the nonpoor females are more disadvantaged with respect to the males than are the poor females. We interpret this to mean that the traditional social constraints of Hispanic culture on girls and women are more vigorous in the middle economic classes.

The sex differences in performance observed for males and females in the Profile sample of young people ages 15 to 22 are primarily in areas that are sex-typed in this society. Whether the sex-typing of particular intellectual and occupational areas as masculine or feminine has evolved from

intuitive wisdom about the patterning of abilities of males and females is not clear from the available literature. Mullins (1980) would argue that the evolution of sex-typed occupational areas is the result of social, political, and economic considerations, and is in only a small way related to biology.

There appears, however, to be evidence of a biological basis of spatial visualizing ability differences between the sexes. High spatial visualizing ability would certainly facilitate performance in a number of the areas in which the males and the females in this sample vary. But as Petersen (1979) points out, even if there is a biological factor in visual-spatial ability, the sex differences in this trait could not alone directly produce the large differences between males and females which appear in course selection in high school and college, and occupational choices in adult life. Positive feedback effects must surely be involved.

The observed sex-related differences in vocational aptitudes must be viewed as the result of a complex interplay of biological factors and social factors. Since experiential factors affect the development of ability, the changing experience of females may result in a pattern of skills among males and females very different from that observed in this study. This may occur in the home, as modes of socialization change; in the school, as teachers and counselors become sex-blind in their treatment of boys and girls with regard to course taking and expectations; and in vocational training, as women begin to realize the importance of developing technical skills for careers in which women have not traditionally participated.

There are obvious vocational implications of these findings. Because of their greater verbal fluency and speed in repetitive tasks, women have a margin of advantage over men in such areas as instruction of children, translation, redaction, typing, clerical work and clerical machine operation,



computer terminal operation, small parts assembly, etc. Because so many women consider only these types of jobs, the supply of labor in these areas is abundant and wages are relatively low. To have access to job markets where supply is smaller and wages higher, women would have to equal or excel men in the capacities they bring to the occupation. Apart from the hard laboring and dangerous jobs that command high wages, the professions are the most attractive alternative for those women who are scholastically successful. Professions like teaching, nursing, and, increasingly, law, medicine and management, are entirely compatible with the profiles of attainment we see for females on the ASVAB tests. But the technical professions, engineering and science, find women at an overall disadvantage because of lower attainment in quantitative reasoning. If equality of opportunity is to be realized in these areas, the sociocultural factors that contribute to the disparities in quantitative skills observed between the sexes must be identified and eliminated. Young women must not only have the opportunity to train themselves in mathematics, they must be actively encouraged to do so. Lack of mathematical preparation is one of the main barriers women must overcome if they are to enter the high-paying technical professions in greater numbers.

## CHAPTER 8

### SUMMARY AND CONCLUSIONS

In the spring and summer of 1980, under the auspices of the Department of Defense, The National Opinion Research Center of the University of Chicago administered the Department of Defense enlistment test--the Armed Services Vocational Aptitude Battery (ASVAB)--to a national probability sample of approximately 12,000 young men and women between the ages of 16 and 23. This battery of tests, in conjunction with other entry standards, is currently used to select personnel for the Army, Navy, Marine Corps, and Air Force, and to assign recruits into appropriate military occupational specialties. It currently consists of ten tests measuring knowledge and skill in the following areas: Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, Numerical Operations, Coding Speed, General Science, Mathematics Knowledge, Mechanical Comprehension, Electronics Information, and Auto and Shop Information. The first four tests measure general trainability, and the second six measure more specialized knowledge relevant to technical vocations. ←

The ASVAB was administered to this sample in order to obtain up-to-date national percentile norms for the tests and to provide data for an assessment of the vocational preparation of the contemporary youth population. Because the subjects used in this study had previously been interviewed as part of the Department of Labor's National Longitudinal Survey of Youth Labor Force Behavior, considerable background information in addition to the aptitude profile information was available on each subject. This study, combining test-score and background data, called the "Profile of American Youth," provided an almost unique opportunity to investigate the vocational potential of a truly representative sample of American young people. The present monograph reports one such examination of these data.

We examined the effects of selected background factors on the ASVAB tests by computing scaled estimates of ability for each subject in the study and subjecting these scores to multivariate analysis of variance based on the background classification. In all tests the scale was chosen so that the mean of the national sample was 500 and the standard deviation 100. The analysis of variance showed that average scores on some or all of the ASVAB tests were related to the background factors--including sex, highest grade completed, sociocultural group (White, Black, Hispanic), economic status (above or below the OMB-defined poverty line), region of residence at age 14, and mother's highest grade completed. (Highest grade completed for the purposes of this analysis was categorized as 0-8 years, some high school (9-11), completed high school (12), some college (13+), and in the case of mother's education, the additional level--completed college.)

In addition, specific relationships were found involving the simultaneous effect of more than one factor. Average score differences between sociocultural group and economic status were significantly affected by highest grade completed, by the sex classification and by the region classification (Northeast, Southeast, Midwest, and West). Average score differences between the sexes were affected by highest grade completed, and to a lesser extent by mother's highest grade completed and sociocultural group. The nature and interpretation of these varied relationships among the background factors and the ASVAB scale-score averages are summarized in this chapter.

#### 8.1 The Sociocultural Groups Differ in ASVAB Test Performance According to Years of Education, Economic Status and Type of Test

The analysis of the Profile data reveals that sociocultural group differences vary depending upon the educational level and economic status of the subjects, and upon the particular test employed. This means that the

interpretation of the observed differences in the test performance of Whites, Blacks and Hispanics must be qualified in respect to the educational, economic and vocational context in which they appear. There are, however, a number of generalizations that can be made that simplify this potentially complex picture.

8.1.1 Educational Effects are Strongest for Tests That Measure School-Intensive Knowledge and Skills

At all education levels, the ordering of the average scores for the three sociocultural groups is similar for most tests: the White groups show the highest scores, the Hispanics are intermediate but somewhat closer to the Whites, and the Black groups are lowest. These results are displayed graphically in Figure 5.1a and b, pp. 132-133. The differences between Hispanics and Whites depend, however, upon the economic classification and the type of test. For tests such as Word Knowledge, Paragraph Comprehension, Numerical Operations, Coding Speed and Mathematics Knowledge, the performance of poor Hispanics is not very different from that of poor Whites, especially at the lower education levels. Among the non-poor, in contrast, Whites clearly perform better than Hispanics at all education levels except 0-8 years, where on the whole there is little to distinguish the two groups.

Blacks score consistently below Whites and Hispanics in all ten of the tests and at all grade levels and economic classes. For tests such as Word Knowledge, Paragraph Comprehension, Numerical Operations, and Coding Speed, the differences between the means for Blacks and those of the other two groups are roughly the same at all education levels, averaging about 50 to 80 score points. For the tests of technical and quantitative knowledge, such as Arithmetic Reasoning, Mathematics Knowledge, Electronics Information, and to a lesser extent General Science and Mechanical Comprehension, the differences

between Blacks and Whites increase with increasing educational levels. To a somewhat lesser extent this is also true of the Hispanics, especially the poor Hispanics. The differences between sociocultural groups are largest for subjects with some college education.

The conclusion from these observations would seem to be that in topics that must be learned primarily in school, the White and the nonpoor groups, who more often have the benefit of schools that have better instructional resources and maintain higher standards, have an advantage over minority groups and the poor. This phenomenon is much less apparent in topics that are less tied to schooling and more part of daily experience in and out of school, such as Word Knowledge, Paragraph Comprehension, Numerical Operations and Coding Speed. In these more basic competencies, the differences between the sociocultural groups are about the same at all education levels. The results discussed in Chapter 5 suggest that access to higher quality education has relatively less effect on these very general types of skills.

#### 8.1.2 Poor and Nonpoor Show Similar Performance at the 13+ Education Level Because of an Artifact in the OMB Definition of Poverty

Both the general and specialized tests show the poor to score below the nonpoor at 0-8, 9-11, and 12 grades completed. But at 13+ (some college), the difference between the poor and the nonpoor largely disappears among Whites and Hispanics. This phenomenon is a consequence of the Office of Management and Budget (OMB) definition of poverty, which classifies self-supporting college students as poor even when the family background is one of comparative affluence. The official definition does not truly reflect the socioeconomic origins of the Whites at this education level. Interestingly, the same is not true of Blacks, and for the technical subject matter is somewhat less true of Hispanics. The likely explanation of this difference is

that more of the Blacks and some of the Hispanics who are in college are living at home and are dependent on parents, and are therefore correctly classified in terms of the economic status of their families. Thus the interaction apparent in these data between education level and economic status is a methodological artifact and not of substantive interest.

8.1.3 English Language Limitations of Hispanics Are Not Very Apparent in the Profile Data

Among Hispanics, the data suggest some effect of English language competency, but the size of the effect is surprisingly small. For the nonlanguage tests in the battery--Numerical Operations and Coding Speed--Hispanics and Whites are very similar in performance, and the largest differences are not, as might be expected, in the language tests--Word Knowledge and Paragraph Comprehension--but in the specialized knowledge tests and Arithmetic Reasoning. Thus, the difference in performance between the Whites and the Hispanics would seem to be one of educational emphasis rather than one of English language competence. On the whole the results give a rather optimistic picture of the vocational preparation of Hispanics relative to Whites at both economic levels. Such differences as exist in the test performance of Whites and Hispanics seem to reflect educational emphasis in the Hispanic communities, and they can be expected to change in a favorable direction as trends in education and acculturation continue.

8.1.4 Some of the Test Scores Reflect Prevailing Practices in Mathematics Teaching in American Schools

Yet another phenomenon is seen in those tests where formal mathematics instruction plays a prominent role--namely, Arithmetic Reasoning, Mathematical Knowledge and to some extent Numerical Operations at lower education levels. For these tests there is a reduced difference between scores of persons who have some high school education compared to those who have completed high school.

As we have pointed out in the discussion of Figure 5.1 on pp. 140-141, this tends to produce a "step" in the plots of average scores vs. education level that is more apparent in these tests than in the tests less dependent on mathematics instruction. We interpret this to be the result of many high school students ending their formal mathematical training at the first and second year of high school, so that less gain is seen in these tests between groups that have completed high school and those that have not. Those students who are college bound, on the other hand, are more likely to continue their mathematics training into the third and fourth years of high school and, therefore, show relatively higher performance on these types of tests both before matriculation and subsequently as they take additional mathematics courses in the early years of college. This observation is consistent with curricular practice in the United States, which treats mathematics beyond grade 10 as largely specialized to the college-bound student. As the Armed Forces require increasingly more technical personnel, these differences in curricular practice may have recruiting implications.

#### 8.1.3 There Are Small But Readily Interpretable Regional Differences in ASVAB Test Performance

The overall cultural group differences discussed above conceal some interesting differences with respect to the four regions of the United States studied--Northeast (bounded by Maryland and Delaware on the south and Pennsylvania on the west and including Puerto Rico and the Virgin Islands), Southeast (bounded by Virginia and West Virginia on the north and Louisiana, Arkansas and Kentucky on the west), Midwest (bounded by Oklahoma on the south and the Dakotas, Nebraska and Kansas on the west), and West (including all remaining states, the Pacific Trust territories and the Canal Zone). These were the regions into which subjects in this study were classified according

to their place of residence at age 14. This classification is intended to represent the major regional influence on the subject's development, but obviously does so only for those subjects who lived in those regions for an appreciable length of time up to age 14.

8.1.6 Blacks and Whites in the Northeast Perform Slightly Better Than Their Counterparts Elsewhere in the Country

Regional effects on the test scores are small, seldom more than 20 scale points, but are statistically significant. The patterns of regional attainment for Blacks and Whites are rather similar. Performance of Blacks and Whites of either economic status tend to be higher in the Northeast, next highest in the Midwest and West, and lowest in the Southeast. But the Auto and Shop Information test is an exception. For this test both the White group and the nonpoor Blacks show progressively higher scores from Northeast to Southeast to Midwest to West, but poor Blacks are lowest in the Midwest. The pattern suggests a perhaps greater involvement of young persons with automobiles and mechanical work as one goes from the East to West across the nation, a trend which roughly parallels extent of automobile use.

The Northeast, as noted, shows higher scores especially in the more academic subjects such as General Science, Arithmetic Reasoning and Mathematics Knowledge, perhaps reflecting higher scholastic standards in these areas in that region. The West is especially strong both in Auto and Shop Information and in Mechanical Comprehension, suggesting greater emphasis on practical skills. The Southeast region, both for Whites and Blacks, generally shows the lowest level of performance, although for Blacks the difference between the Southeast and West is small. Blacks generally perform better in the Northeast and Midwest.



8.1.7 The Test Performance of Hispanics is Highest in the Southeast

A quite converse pattern is seen for the Hispanics. The highest levels of performance are in the Southeast and Midwest with both the East and West showing generally lower levels. In fact, the nonpoor Hispanics of the Southeast and Midwest perform at about the same level as the nonpoor Whites in Paragraph Comprehension, Numerical Operations and Coding Speed. In General Science, Arithmetic Reasoning, and Word Knowledge Hispanics of the Southeast perform almost at the same level as the nonpoor Whites. Some of the largest differences between the nonpoor and poor classes occur among the Hispanics, especially those of the Midwest, suggesting that economic factors are more dominant for Hispanics than for either Whites or Blacks.

These effects are easily understood in terms of the origins of the Hispanic populations in these four regions of the country. Hispanics of the Northeast are drawn mainly from lower economic level Puerto Ricans who now reside in the large metropolitan areas of the Northeast in close proximity to the Black populations. Both nonpoor and poor Hispanics are performing at about the same level as Blacks in the Northeast (which is the region of the country where Blacks show their highest performance). In fact, the performance of poor Hispanics in the Northeast is below that of poor Blacks in that region. The Hispanics represented in this sample in the Southeast are largely the children of middle-class Cubans who came to Miami when Castro came to power in the 1950s. They have been largely assimilated into the larger social-economic community in Florida and their performance differs little from that of members of the majority culture in that region. In the Midwest, the higher income Hispanics are well assimilated residents of long standing in the major areas such as Chicago and are performing only slightly below the level of the majority culture. Poor Hispanics in the Midwest, however, tend to be

new residents and are performing at a level closer to the range of the Blacks. In the West, both nonpoor and poor Hispanics represent persons who have close connections with Mexico and the Spanish language community and perform on these tests at a generally lower level, midway between Whites and Blacks. Moreover, there are smaller differences between nonpoor and poor Hispanics in the West. All in all, these results, especially the regional effects among Hispanics, tend to demonstrate the impact of the cultural background on performance on these types of vocational tests.

8.1.8 Some Blacks Are Performing at High Levels, But Their Relative Numbers Are Small

Neither the effects of education, nor those of economic status, nor those of region are very revealing of the ultimate origins of the differences in vocational test performance among the sociocultural groups. Blacks, for example, show many education and region effects similar to those of Whites and Hispanics, but their average scores are almost always lower, in some cases by as much as 100 scale points. When the score distributions (shown in Figures 5.1a, 5.1b, 5.2; pp. 132-133, p. 149) are examined in detail, however, there is some evidence of a distinguishable small group of Blacks, amounting to perhaps 10 percent of the Black population, whose scores are comparable to those of Whites. We suppose these Blacks to be those who are participating fully in the majority culture as represented in the content of the vocational tests in this study. Unfortunately, the sources of these high achieving Blacks, who are not accounted for by educational level, cannot be more precisely identified in the present data. For this reason, we have had to rely on studies in the literature for some insight into the reasons for the greater score differentiation seen among Blacks compared to Whites of similar education and economic class.

8.1.9 Evidence from Recent Studies Contradicts the Theory That Differences in the Average Test Performance of the Sociocultural Groups Are of Genetic Origin

One unsubstantiated theory that appears from time to time attributes the disparity in cognitive test performance between Blacks and Whites to genetic differences between the subpopulations. It is supposed that there exist genes, as yet unidentified, that are critical for cognitive functioning and, owing to past selective forces, are less widely represented among Blacks than among Whites. The claim is that the substantial differences between mean cognitive test scores of Whites and Blacks in the United States arise from this source. These theories are based, however, on the erroneous assumption that the moderately high levels of heritability that have been found for certain general and special cognitive abilities also account for differences in the average performance levels in socially defined populations.

The clearest evidence that cognitive functioning is influenced by biological inheritance is that identical twins, even those reared separately from near birth onwards, are consistently found to be more similar in levels of cognitive test performance than fraternal twins, and that the performance of early adopted children resembles more their natural parents than their adoptive parents. The bulk of evidence from these types of family study supports the conclusion that genetic effects account for an appreciable part of the individual differences in performance that one sees within populations. Such studies are, however, always carried out within relatively homogeneous social groups, and do not attempt to account for differences between the mean levels of populations.

On the contrary, studies of other quantitative traits, especially those having to do with body size and proportion, show that the average levels of equally heritable physiological traits can change by a comparable amount in

only two generations. Mature height, for example, has almost exactly the same degree of heritability as I.Q. test performance. But in response to better nutritional environments and health care, height has been increasing at about 3.0 centimeters per generation in developed and developing countries.

Increases of this magnitude have been occurring in Europe and the United States since the middle of the last century, and have occurred at an even greater rate in Japan and other Far Eastern countries since World War II. Most of the gain is due to increases in leg length, with the result that mature body proportions are also changing. Thus, the stocky build traditionally associated with the Japanese turns out to be an environmental rather than a genetic difference between that population and Western populations. These counterexamples firmly establish the fact that within population heritabilities of quantitative traits, even as high as are found for height or cognitive test performance, do not imply that the differences between the average levels of the trait in socially distinct populations are of genetic origin.

Studies that separate the effects of biological and social inheritance on behavioral indices are difficult to carry out because persons who are more closely related biologically tend to occupy more similar social environments than those who are distantly related or unrelated. As a result, the effects of heredity and environment on cognitive development are thoroughly confounded in most studies. One approach to separating these effects is to examine the relationship between the behavioral measures and biological indicators of genetic similarity. As we have discussed in Chapter 5, p. 164, one such study, carried out by Sandra Scarr and her colleagues, made use of the fact that the proportions of genes carried by a given person that arose in the European and African gene pools can be estimated from certain minor blood

group markers that have widely different frequencies in these populations. If the putative genes for cognitive ability also have different frequencies in these populations, then cognitive test scores should correlate with the proportion of admixture of these gene pools within individuals. Scarr and her colleagues observed no such correlation, indicating that the differences we see in the average test performance of populations in the United States deriving from European and African origins are not substantially of genetic origin.

Another approach to separating biological and environmental effects on cognitive development is to examine the performance of children who as infants have been adopted across sociocultural group or economic class lines. We refer in Chapter 5, pp. 161-165, to studies of this type carried out by Scarr and Weinberg and independently by Elsie G. J. Moore. In both studies, Black children adopted at an early age into White middle-class families were administered standard intelligence tests at several ages during childhood. These children were found to perform at a mean level typical of White children of comparable age, rather than at the lower level typical of such children raised in Black families. Moore's study was especially informative because she compared Black children adopted into middle-class White homes with those adopted into middle-class Black homes. She found that the children in the White homes scored on the tests at a mean level typical of White middle-class children and those in the Black homes scored at a mean level typical of Black middle-class children. There was nothing in the backgrounds of these children to suggest that the observed differences could be attributed to selective placement by the adoption agency. Rather, the difference appeared to be due to the cumulative effect on children of being reared in homes and neighborhoods with different exposure to and emphasis on the kinds of knowledge,

skills and task orientations that are required for successful performance on certain cognitive tests.

The conclusion from the Scarr and Moore studies is that when Black children are adopted into White middle-class homes at an early age and raised completely within that environment, their performance on cognitive tests is indistinguishable from that of White children of the same age. The similarity of this effect of the sociocultural environment on behavioral development with that of improved diet and medical care on physiological growth and development is indeed striking. In both cases, variation of the traits within population is substantially heritable, but the cumulative impact of a change in nurturance is nevertheless seen as an increase in the average level of the population in a measure of overall growth, cognitive on the one hand and physical on the other. These adoption studies, and the evidence of intergenerational increase in stature in developing countries, suggest a resolution to the longstanding "nature-nurture" controversy in the study of human behavior. Both influences act simultaneously; but whereas hereditary effects are easily seen in individual differences, environmental effects are more clearly expressed in differences in the average levels of traits between generations and between populations.

#### 8.1.10 A Community Norm Theory Seems to Be the Best Explanation of the Sociocultural Group Differences

Although the studies cited above demonstrate the effect of the sociocultural environment on cognitive test performance, they do not give any indication of the processes by which these effects occur. In our review of the literature in Section 5.4, however, we cite several theories of how such processes operate. One such theory is contained in the formal mathematical models of status attainment that attempt to account for variation in general

measures of school achievement similar to the ASVAB tests. The explanatory (exogenous) variables in such models typically include father's occupation, income and social status, and the child's general ability as indexed by an IQ score. These variables are supposed to affect the child's ambition for attainment and ultimately his or her other school achievement. As causal explanation, however, these models are questionable. Because cognitive measures such as IQ scores are, as we have seen, influenced by the sociocultural group and economic class in which the child is raised, these models would seem to describe relationships among effects rather than responses to ultimate causes. Thus, the formal status attainment models do not seem to offer the type of explanation we are seeking.

Neither do the other prevailing theories of social and economic group differences--namely, the linguistic minority and early deprivation theories--seem adequate to explain the results of the Profile Study. Deficits due to lack of proficiency in the majority language do not account for differences in the non-language tests of the ASVAB, nor in the pattern of results of the other tests in light of their language content. Similarly, effects of early deprivation do not explain the increasing differences between Blacks and Whites at increasing education levels through some college.

A more satisfactory explanation is simply that the communities represented by the present more-or-less exclusive sociocultural subpopulations in the United States maintain, for historical reasons, different norms, standards and expectations concerning performance within the family, in school and in other institutions that shape children's behavior. Young people adapt to these norms and apply their talents and energies accordingly. These norms are so much a part of the fabric of the face-to-face community setting in which family, relatives, friends and acquaintances interact that the standards

of the larger society--represented here by norms of test performance--cannot easily displace or alter them. Although the task of raising standards of test performance is traditionally assigned in the abstract to schools, actual schools in the United States are responsible first to their communities, only secondarily to states, and only indirectly to the nation. Thus, norms of the immediate community tend to be perpetuated, even when the result is the range of group differences such as that seen in the Profile study.

There is ample reason to believe, however, that many schools could work toward a higher standard than that which prevails. But the process of raising standards involves the whole of the communities that these schools serve, and must extend to parents, families, social groups, and every institution in which members of the community participate. Moreover, the community norm is part of the cultural transmission from one generation to another, and the process must work on the scale of decades, not just years. Only steady and consistent efforts over an extended period can produce significant changes in the community norm and the educational standards it determines.

#### 8.2 Mother's Education and Years of Formal Schooling Are Strong Predictors of Vocational Test Performance

Mother's education (Mother's Highest Grade Completed), discussed in Chapter 6, differs from the other factors investigated in this study in that it is largely a within-family influence on the performance of the children. It stands apart from the broad social variables of educational level, socio-cultural group, economic status, and region of the country, which act much less directly. This is perhaps the reason that mother's education, unlike the other background factors, shows only weak interaction with any of the other background factors.



8.2.1 Mother's Education Is Strongly Associated with Better Test Performance, But Interacts Very Little with Other Background Factors

Mother's education shows a strong and direct association with test performance, especially in those tests which depend upon language and instruction (Word Knowledge, Paragraph Comprehension, General Science, Arithmetic Reasoning, and Mathematics Knowledge). As might be expected, the tests that are less related to formal learning and language depend much less strongly on mother's education. These include Coding Speed, Auto and Shop Information, Mechanical Comprehension and Electronics Information. The mother's educational variable is of interest primarily in the extent to which it shows how family background characteristics are reflected in test performance. For Word Knowledge, for example, the difference in average scores between children of mothers with only 8 or less years of education versus those with college degrees is about 60 scale points. This difference is not as great as the direct effect of the young person's own education, which amounts to nearly 150 scale points, but it indicates the close association between the intellectual attainments of the mother and those of her children. In large-scale studies where a wide range of education levels is represented, mother's education is an effective indicator of the family's contribution to children's attainment.

8.2.2 Among Persons of Different Ages Who Have the Same Number of Years of Education, Some Test Scores Increase with Age, Some Decline, But for the Most Part Levels of Performance Remain About What They Were at the End of Formal Education

Although age and highest grade completed tend to be confounded in the Profile data, there is enough variation in age within the highest grade completed groups to enable us to draw some conclusions about age effects independent of schooling. The supplementary analysis of the Profile data discussed in Section 6.2 reveals that performance on tests such as Arithmetic Reasoning

and Numerical Operations, representing intensive, school-learned skills, tends to decline after a person completes formal schooling. Performance on test content that is learned mostly by experience, such as Auto and Shop Information and Electronics Information, tends to improve after leaving school.

These effects are small, however, compared with the general tendency for the level of performance on all tests to remain near their levels at the end of schooling. The analysis confirms the impact of formal school completion in channeling the vocational opportunities for American youth.

### 8.3 Sex Differences Vary Most Strongly with the Type of Test

As we have shown in Chapter 7, the size, and in some cases the direction, of the differences between males and females vary greatly from one ASVAB test to another. In particular, men and women have distinctly different profiles of both general ability and special knowledge. Under present conditions of education, women are at a competitive disadvantage relative to men in tests requiring technical knowledge and quantitative skills; they have an advantage in tests requiring fluent and accurate information processing and use of language. Studies in the educational, psychological, and physiological literature suggest that these differences result partly from sexual specialization of education and experience, and partly from biologically intrinsic differences.

#### 8.3.1 Men and Women Differ Most in Knowledge of Topics That Are Traditionally Specialized by Sex, But the Effect Is More Apparent in Whites and Hispanics Than Blacks

Typical of the sex differences where average scores of men exceed those of women is performance on the Auto and Shop Information test. Among nonpoor Whites and Hispanics, the average scores of male subjects exceed those

of female subjects by slightly more than 100 scale points. The sex differences for Blacks of either economic status are considerably smaller for this test, amounting to not more than 50 scale points. They are also smaller for poor persons in all subgroups.

This pattern of attainment suggests that with greater access to automobile and shop experience, it is the young men who benefit most. By circumstance or choice, women do not participate in study or activities that bring them into contact with this kind of information. Among Blacks, poor Whites and Hispanics, auto and shop information is less available or less exploited by both males and females and the sex differences are smaller.

Similar patterns of greater male attainment appear in the Mechanical Comprehension and Electronics Information tests but are not as pronounced as in Auto and Shop Information. The largest sex difference for these tests is about 75 scale points. Still smaller but significant sex differences in favor of males are also seen in the General Science and Arithmetic Reasoning tests, but the interactions with sociocultural group and economic status are diminished. Although Blacks show larger sex differences on General Science and Arithmetic Reasoning than is the case with Auto and Shop, Mechanical Comprehension and Electronics Information, these differences are still less than the sex differences shown by Whites on the same tests.

Neither the Word Knowledge nor the Mathematics Knowledge test shows consistent or statistically significant sex differences.

#### 8.3.2 Women Excel in Reading Comprehension and in Tasks Requiring Fast and Accurate Responses

Patterns of female advantage are seen in Coding Speed, Numerical Operations, and Paragraph Comprehension. In Coding Speed, females exceed males by about 50 scale points, with the size of the difference slightly greater among

Whites than among Hispanics or Blacks. A similar pattern is seen in the Numerical Operations test, but the sex difference is reduced to about 25 scale points. However, very little sex difference is apparent on this test for poor Blacks, even though nonpoor Blacks show the same 25 scale-point advantage of females over males as do the other groups. Females also score higher on Paragraph Comprehension, but the sex differences favoring females are even smaller and are virtually absent for poor Blacks. The results suggest that the cultural and educational disadvantages that impair the test performance of poor Blacks fall more heavily on females relative to males than is true of nonpoor Blacks or the other groups.

Female performance on Numerical Operations, Coding Speed, and Paragraph Comprehension may be explained by psychological research that shows women to be uniformly superior to men in so-called "fluent-production" tasks. These are tasks in which an overlearned response must be made quickly and accurately. Numerical Operations is a speeded test in which the subject is required to perform as many simple 1-digit computations as possible in a fixed length of time. Coding Speed is a similar test in which the subject must look up code numbers of words in a table and mark them on the answer sheet. The Paragraph Comprehension test is a measure of basic reading comprehension. Although the time allotted to this test is not especially short, an element of reading fluency may also be involved that helps account for the sex difference. It has been suggested that women react more positively than men to these types of repetitive tasks and thus perform better. But the few studies that have investigated this phenomenon are far from definitive.

The present study confirms that a pattern of male advantage for quantitative and technical tasks and female advantage for fluency tasks, which has

heretofore been seen in studies of White secondary school and college students, is general to all sociocultural groups and economic levels. It is especially significant that the Arithmetic Reasoning test shows sex differences in favor of males while the Mathematics Knowledge test does not uniformly do so. This suggests that Arithmetic Reasoning, which requires the solution of "word problems" using numerical quantities, is in some important respect different from the Mathematics Knowledge test, which asks only for facts about mathematics and the execution of formal algebraic rules. There is a substantial body of evidence that males consistently exceed females in the ability to visualize relationships and movements among two- and three-dimensional figures and objects. It has been conjectured that the solution of arithmetic word problems is facilitated by visualizing the quantities and relationships involved. The contrasting sex differences of the Arithmetic Reasoning and Mathematics Knowledge tests in this study support that conjecture.

As we discuss in Chapter 7, there is also considerable evidence in the literature that the difference in male and female ability to visualize spatial relations is influenced by physiological differences between the sexes. In particular, males in whom the normal hormonal changes of adolescence are pathologically absent have levels of spatial ability similar to those of females. The connection between spatial ability and quantitative and configural problem solving is a possible source of the sex differences seen in the Arithmetic Reasoning and Mechanical Comprehension tests.

### 8.3.3 Some Sex Differences in Test Performance Increase with Education; Others Decrease

When the sex differences are examined as a function of the subjects' highest grade completed, it is found that for Auto and Shop Information,

Mechanical Comprehension, Electronics Information, General Science, and Arithmetic Reasoning, the differences between the sexes increase with increasing education. This is interpreted as another expression of the principle seen in the interaction of sex differences with sociocultural group and economic status for Auto and Shop Information--namely, that where there is greater opportunity for exposure and experience in those areas where males generally perform better than females, males benefit more from such exposure and thus increase the disparity of performance between the sexes. With the increasing specialization that comes with education, men and women also become more specialized in sex-typed knowledge and skills that lead to different profiles of average vocational test scores. At lower educational levels, experience and instruction are more common between the sexes and sex differences in test performance are not as large.

In the fluent production tasks (measured by the Numerical Operations and Coding Speed tests) where females score higher, sex differences remain almost constant as a function of level of education. This suggests that sex differences in fluency are due to differences between men and women which are not much influenced by education. In Paragraph Comprehension, the sex difference favoring females remains about the same until subjects have completed 13 or more years of schooling. At this point, both men and women have mastered simple reading tasks and the test no longer differentiates between the sexes. The Word Knowledge test shows no sex difference at any grade level, and the Mathematics Knowledge test shows a sex difference in favor of males only at 13+ years, where the greater tendency of men to take mathematics courses in college finally has an effect. This is in contrast with Arithmetic Reasoning, which shows steadily increasing sex differences in favor of males from the 9th grade on.

#### 8.3.4 Differences in the Vocational Test Performance of Men and Women Affect Their Vocational Opportunities

The implications of these patterns of sex differences for vocational choice are well known and widely discussed. The disadvantage to females which results from their lower performance in technical areas such as Auto and Shop Information, Mechanical Comprehension, Electronics Information and General Science is reflected in the relatively small numbers of women who enter the technical fields. The parity between men and women with respect to Word Knowledge and Paragraph Comprehension is reflected in the relative equality of men and women in the literary and writing fields. And finally, the rather marked advantage of women in the fluent production tasks, which are so much a part of clerical work such as business machine operation and word processing, suggests one of the reasons why women largely displace men in these fields. The present study gives a more complete account of these vocationally relevant sex differences than has previously been available.

#### 8.4 Differences in Test Performance of Young People at the Same Level of Education in Different Social and Economic Groups Show the Need for More Comprehensive Standards of Educational Attainment

The Profile data reveal striking disparities between sociocultural and economic groups on all tests. These disparities between groups exist even when persons of the same educational level are compared. In Chapter 5 we argued for differences in community attainment standards as the proper explanation of group test scores differences. Some sociocultural groups are more conscious than others of the main avenues of vocational opportunity, and of the kinds of preparation needed to take advantage of them. They know the importance of language and mathematics skills, and of familiarity with the scientific and technical concepts on which so much of our economic life depends. Young people from communities with a higher achievement standard have a marked advantage on tests designed to predict occupational success.

Everyone would benefit from the diffusion of high achievement standards. The disparity between social groups would be reduced and the nation's human resources would increase. The social and economic cost of personal failures would decrease, and more skill and talent would be available for every aspect of craftsmanship, technology, art, and science. To believe that levels of attainment in diverse social groups can be influenced by a shared commitment to higher standards is far from utopian. At the end of Chapter 5 we cited the evidence from the school achievement surveys of the National Assessment of Educational Progress to show that differences between Blacks and Whites at ages 9 and 13 lessened during the period of educational restructuring in the 1970s. This trend is unmistakable and applies to almost all of the important school subjects at these age levels. If it continues through the 1980s and 1990s, we can look forward to much reduced differences between advantaged and disadvantaged young people at the end of this century. Quite possibly, smaller differences will then prevail between the sexes as well as between sociocultural groups. Moreover, if the high standards which presently exist for some are diffused to all, this society will benefit from future generations that have broader potential for productive and creative lives than any that have gone before.



REFERENCES

- Aguirre, B. Differential migration of Cuban social races. Latin American Research Review, 1976, 11, 13-124.
- Allen, W.R. Preludes to attainment: Race, sex, and student achievement orientations. The Sociological Quarterly, 1980, 21, 65-79.
- Alvarez, R. The psycho-historical and socioeconomic development of the Chicano community in the United States. In C. A. Hernandez, M. J. Haug, and N. Wagner (Eds.), Chicanos: Social and Psychological Perspectives. St. Louis, MO: C. V. Mosby Co., 1976.
- Anastasi, A. Psychological Testing. New York: MacMillan Publishing Co., Inc., 1982.
- Anastasi, A., and Cordova, F.A. Some effects of bilingualism upon intelligence test performance of Puerto Rican children in New York City. Journal of Educational Psychology, 1953, 44, 1-19.
- Armstrong, J.M. Achievement and Participation of Women in Mathematics: An Overview. Report of a two-year study funded by the National Institute of Education, March, 1980.
- Armstrong, J.M. Achievement and participation of women in mathematics: Results of two national surveys. Journal for Research in Mathematics Education, 1981, 12 (5), 356-372.
- Asbury, C.A. Cognitive correlates of discrepant achievement in reading. Journal of Negro Education, 1973, 42 (2), 123-33.
- Ashton, G.C., Polovina, J.J., and Vandenburg, S.C. Segregation analysis of family data for 15 tests of cognitive ability. Behavioral Genetics, 1979, 9, 329-347.
- Astin, A.W. Minorities in American Higher Education. San Francisco: Jossey-Bass. 1982.
- Astin, H.S. Sex Differences in Mathematical and Scientific Precocity. In J. C. Stanley, D.P. Keating, and L.H. Fox (Eds.), Mathematical Talent: Discovery, Description, and Development. Baltimore: The Johns Hopkins University Press, 1974.
- ASVAB Working Group. History of the Armed Services Vocational Aptitude Battery: 1974-1980. Washington, D.C.: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics), March, 1980.
- Atwater, D.C., and Abrahams, N.M. Evaluation of Alternative ASVAB Composites for Selected Navy Technical Schools. NPRDC-TR 80-15. San Diego, Navy Personnel Research and Development Center, 1980.

- Baratz, S., and Baratz, J. Early childhood intervention: The social science base of institutional racism. Harvard Educational Review, 1970, 40, 29-50.
- Barral, D.P. Achievement Levels Among Foreign Born and Native Born Mexican American Students. San Francisco: R & E Research Associates, 1977.
- Bayer, A.E. The Black College Freshman: Characteristics and Recent Trends. (Research Report No. 8.) Washington: American Council of Education, 1972.
- Bean, J. P. What's Happening in Mathematics and Science Classrooms: Student-Teacher Interactions. Paper presented at meeting of American Educational Research Association, San Francisco, 1976.
- Bem, S. On the utility of alternative approaches for assessing psychological androgyny. Journal of Consulting and Clinical Psychology, 1977, 45, 196-205.
- Bem, S., and Lenney, E. Sex-typing and the avoidance of cross-sex behavior. Journal of Personality and Social Psychology, 1976, 33 (1), 48-54.
- Benbow, C.P., and Stanley, J.C. Sex differences in mathematical ability: fact or artifact? Science, 1980, 210, 1262-1264.
- Bennett, G.K., Seashore, H.G., and Wesman, A.G. Differential Aptitude Tests. New York: The Psychological Corporation, 1974.
- Bennett, G.K., Seashore, H.G., and Wesman, A.G. Fifth Edition Manual for the Differential Aptitude Tests Forms S and T. New York: The Psychological Corporation, 1974.
- Bereiter, C., and Engelmann, S. Teaching Disadvantaged Children in the Preschool. Englewood Cliffs, NJ: Prentice-Hall, 1966.
- Bernstein, B. Social class and linguistic development. In A.J. Halsey, J. Floud, and C.A. Anderson (Eds.), Education, Economy, and Society. New York: Free Press, 1961.
- Blank, M., and Solomov, F. A tutorial program to develop abstract thinking in socially disadvantaged preschool children. Child Development, 1968, 39, 379-389.
- Blau, P.M., and Duncan, O.D. The American Occupational Structure. New York: Wiley, 1967.
- Blau, Z.S. Black Children-White Children: Competence, Socialization, and Social Structure. New York: Free Press, 1981.
- Biau, Z.S. Exposure to child-rearing experts: A structural interpretation of class-color differences. American Journal of Sociology, 1964, 64, 597-608.
- Bloom, B.S. Stability and Change in Human Characteristics. New York: Wiley, 1964.

- Bloom, B.S., Davis, A., and Hess, R. Compensatory Education for Cultural Deprivation. New York: Holt, 1963.
- Bock, R.D. Basic issues in the measurement of change. In D.N.M. de Gruijter and L.J.T. van der Kamp (Eds.), Advances in Psychological and Educational Measurement. London: Wiley and Sons, 1976.
- Bock, R.D. Multivariate Statistical Methods in Behavioral Research. New York: McGraw-Hill, 1975.
- Bock, R.D. Review of Educability and Group Differences, by A.R. Jensen. Perspectives in Biology and Medicine, 1974, 17, 594-597.
- Bock, R.D. Word and image: Sources of the verbal and spatial factors in mental test scores. Psychometrika, 1973, 38, 437-457.
- Bock, R.D., and Haggard, E.A. The use of multivariate analysis of variance in behavioral research. In D. Whitla (Ed.), Handbook of Measurement and Assessment in Behavioral Sciences. Reading, MA: Addison-Wesley, 1968.
- Bock, R.D., and Kolakowski, D. Further evidence of sex-linked major-gene influence on human spatial visualizing ability. American Journal of Human Genetics, 1973, 25, 1-14.
- Bock, R.D., and Mislevy, R.J. The Profile of American Youth: Data Quality and Analysis of the Armed Services Vocational Aptitude Battery. Chicago: National Opinion Research Center, 1981.
- Bock, R.D., and Perline, R. A Lod score method for detecting linkage on the X-chromosome between a marker locus and a major-gene locus for a qualitative character. Behavior Genetics, 1979, 9 (3), 139-149.
- Bock, R.D., and Thissen, D. Familial resemblance in patterns of growth in stature. In W.E. Nance (Ed.), Twin Research: Psychology and Methodology. New York: Alan R. Liss, 1978.
- Bodmer, W.F., and Cavalli-Sforza, L.L. The Genetics of Human Population. San Francisco: Freeman, 1971.
- Bouchard, T.J., Jr., and McGee, M.G. Sex differences in human spatial ability: Not an X-linked recessive gene effect. Social Biology, 1977, 24, 332-335.
- Bouchard, T.J., Jr., and McGue, J.L. Familial studies in intelligence: A review. Science, 1981, 212, 1055-1059.
- Bray, D.W., and Moses, J.L. Personnel selection. In P.H. Mussen and M.R. Rosenzweig (Eds.), Annual Review of Psychology, 1972, 23, 545-576.
- Brim, O.G., and Kagan, J. (Eds.). Constancy and change in human development. Cambridge, MA: Harvard University Press, 1980.
- Bronowski, J. Mathematics. In D. Thompson and J. Reeves (Eds.), The Quality of Education. London: Muller, 1947.

- Brookover, W.B., Beady, C., Flood, P., Schwietzer, J., and Wisenbaker, J. School, Social Systems, and Student Achievement: Schools Can Make a Difference. New York: Praeger, 1979.
- Brookover, W.B., and Erickson, E.L. Society, Schools and Learning. Boston: Allyn & Bacon, 1969.
- Broverman, D.M., and Klaiber, E.L. Negative relationships between abilities. Psychometrika, 1969, 34, 5-20.
- Brown, F., Carter, D.G., and Harris, J.J. Minority students, ability grouping, and career development. Journal of Black Studies, 1978, 8, 477-488.
- Brown, G.H., Rosen, N.L., Hill, S.T., and Olivas, M.A. The Condition of Education for Hispanic Americans. Washington, D.C.: National Center for Education Statistics, 1980.
- Brush, L.R. Encouraging Girls in Mathematics: The Problem and the Solutions. Cambridge, Mass.: Abt Associates, 1980.
- Burnett, S.A., Lane, D.M., and Dratt, L.M. Spatial visualization and sex differences in quantitative ability. Intelligence, 1979, 3, 345-354.
- Burton, N.W., and Jones, C.V. Recent trends in achievement levels of Black and White youth. Educational Researcher, 1982, 12, 10-14, 17.
- Carnegie Council on Policy Studies in Higher Education. Giving Youth a Better Chance: Options, for Education, Work, and Service. San Francisco: Jossey-Bass, 1980.
- Carter, T.P. Mexican-Americans in School: A History of Educational Neglect. New York: College Entrance Examination Board, 1970.
- Casserly, P.L. Factors affecting females' participation in advanced placement in mathematics, chemistry, and science. In L.H. Fox, L. Brody, and D. Tobin (Eds.), Women and the Mathematical Mystique. Baltimore: The Johns Hopkins University Press, 1980.
- Cattell, R.B. The theory of fluid and crystallized intelligence. British Journal of Educational Psychology, 1967, 37, 209-224.
- Centra, J.A. Women, Men, and the Doctorate. Princeton, NJ: Educational Testing Service, 1974.
- Chan, K.S., and Rueda, R. Poverty and culture in education: Separate but equal. Exceptional Children, 1979, 45, 422-428.
- Clarke, A.M., and Clarke, A.D.B. Early Experience: Myth and Evidence. New York: Free Press, 1976.
- Cleary, T., Humphreys, L., Kendrick, S., and Wesman, A. Educational uses of tests with disadvantaged students. American Psychologist, 1975, 30, 15-41.
- Cole, M., and Bruner, J. Cultural differences and inferences about psychological processes. American Psychologist, 1971, 26, 867-875.

- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., and York, R.L. Equality of Educational Opportunity. Washington, DC: U.S. Government Printing Office, 1966.
- College Board, Admission Testing Program. National Report, 1981. New York: College Entrance Examination Board, 1981.
- Connor, J.M., and Serbin, L. Behaviorally based masculine-feminine activity preference scales for preschoolers: Correlates with classroom behaviors and cognitive tests. Child Development, 1977, 48, 1411-1416.
- Cook, M.A. and Alexander, K.L. Tracking and training in the academic switchboard: Curricula, course-work and the equity of school organization. (Unpublished paper.) Cleveland: Case Western Reserve University, 1981.
- Cooper, B. An analysis of reading achievement of White and Negro pupils in certain public schools in Georgia. School Review, 1964, 72, 462-471.
- Crandall, V., Preston, A., and Rabson, A. Maternal reactions and the development of independence and achievement behavior. Child Development, 1960, 31, 243-251.
- Crim, A. A community of believers. Daedalus, 1981, 110 (4), 145-162.
- Crim, A. Community of believers. In Educating Our Citizens: The Search for Excellence. Washington, D.C.: Center for National Policy, 1983.
- Crim, A. (Personal communication, 1983.)
- Cresswell, J.L., and Houston, G. A study of sex-related differences in mathematics achievement of Black, Chicano, and Anglo adolescents. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, Massachusetts, April, 1980.
- Cummins, J. Linguistic interdependence and educational development of bilingual children. Review of Educational Research, 1979, 49, 222-251.
- Cuttance, P. Do schools consistently influence the performance of their students? Educational Review, 1980, 32 (3), 267-80.
- Cuttance, P. Psycho-social environment of the school: Reinforcing the social structure. Political Science, 1979, 31 (1), 39-53.
- Dale, P.S. Language development: Function and structure. Hinsdale, IL: Dryden Press, 1976.
- Dave, R.H. The identification and measurement of environmental variables that are related to educational achievement. Unpublished Ph.D. dissertation, Department of Education, University of Chicago, 1963.
- Dawkins, M.P. Educational and occupational goals: Male versus female Black high school seniors, Urban Education, 1980, 15, 231-242.

- DeFries, J.C., Ashton, G.C., Johnson, R.D., Kuse, A.R., McClearn, G.E., Mi, M.P., Rashad, M.N., Vandenberg, S.G., and Wilson, J.R. Parent-offspring resemblance for specific cognitive abilities in two ethnic groups. Nature, 1976, 261, 131-133.
- Deutsch, C. Social class and child development. In B.M. Caldwell and H. Ricciuti (Eds.), Review of Child Development Research, Vol. 3. Chicago: University of Chicago Press, 1973.
- Deutsch, M., Fishman, J., Kogan, L., North, R., and Whiteman, M. Guidelines for testing minority children. Journal of Social Issues, 1964, 20, 129-145.
- Douglas, J.W.B. The Home and the School. London: McGibbon and Kee, 1964.
- Duncan, O.D., Featherman, D.L., and Duncan, B. Socioeconomic Background and Achievement. New York: Seminar Press, 1972.
- Edmunds, R. Effective schools for the urban poor. Educational Leadership, 1979, 37, 15-18.
- Edwards, O. Cohort and sex changes in Black educational attainment. Sociology and Social Issues, 1979, 59 (2), 110-120.
- Eitelberg, M. and Doering, Z.D. Profile in perspective: The policy and research implications of the "Profile of American Youth." Paper presented at the Annual Meeting of the American Psychological Association, Washington, D.C., August 24, 1982.
- Entwisle, D.R. Socialization of language behavior and educability. In M. Maher and W. Stallings (Eds.), Culture, Child, and School. Monterey, CA: Brooks/Cole, 1975.
- Erikson, E.H. Childhood and Society, Second Edition. New York: W.W. Norton, 1963.
- Erikson, E.H. Childhood and Society, Third Edition. New York: W.W. Norton, 1968.
- Erlenmeyer-Kimling, L., and Jarvik, L. Genetics and intelligence. Science, 1963, 142, 1477-1479.
- Erlick, A.C. and LeBold, W.K. Factors Influencing the Science Career Plans of High School Students. West Lafayette, IN: Purdue Opinion Panel, Measurement and Research Center, Purdue University, 1975.
- Erlick, A.C. and Lebold, W.K. Factors Influencing the Science Career Plans of Women and Minorities. West Lafayette, Indiana: Purdue Opinion Panel, Measurement and Research Center, Purdue University, 1977.
- Ernest, J. Mathematics and sex. American Mathematical Monthly, 1976, 83, 595-614.
- Farran, D.C., Haskins, R., and Gallagher, J.J. Poverty and mental retardation: A search for explanations. In J.J. Gallagher (Ed.), New Directions for Exceptional Children, Vol. 1. San Francisco: Jossey-Bass, 1980.

- Featherman, D.L. Schooling and occupational careers: Constancy and change in worldly success. In O. Brim and J. Kagan (Eds.), Constancy and Change in Human Development. Cambridge, MA: Harvard University Press, 1980.
- Featherman, D.L., and Hauser, R.M. Sexual inequalities and socioeconomic achievement in the U.S., 1962-1973. American Sociological Review, 1976, 41, 462-483.
- Feldmann, S. and Weiner, M. A fourth validation of a reading prognosis test for children of varying socioeconomic status. Educational and Psychological Measurement, 1966, 26, 463-470.
- Fennema, E. Influences of selected cognitive, affective, and educational variables on sex-related differences in mathematics learning and studying. In Women and Mathematics Research: Perspectives for Change. Washington, DC: National Institute for Education, 1977.
- Fennema, E. Mathematics learning and the sexes: A review. Journal of Research in Mathematics, 1974, 5, 126-139.
- Fennema, E. Women and mathematics: State of the art review. Paper presented at the Equity in Mathematics Core Conference, Reston, Virginia, February, 1982.
- Fennema, E. The sex factor: Real or not in mathematics education. In E. Fennema (Ed.), Mathematics Education Research: Implications for the 80s. Alexandria, VA: Association for Supervision and Curriculum Development, August, 1981.
- Fennema, E. Teachers and sex bias in mathematics. Mathematics Teacher, 1980, 3, 169-173.
- Fennema, E. and Sherman, J. Sex-related differences in mathematics achievement, spatial visualization, and affective factors. American Educational Research Journal, 1977, 14 (1), 51-71.
- Fernandez, C., Espinosa, R.W., and Dornbusch, S.M. Factors perpetuating the low academic status of Chicano high school students. Research and development memorandum No. 138. Stanford, CA: Stanford Center for Research and Development in Teaching, Stanford University, 1975.
- Finn, J.D. MULTIVARIANCE: Univariate and Multivariate Analysis of Variance, Covariance, Regression, and Repeated Measures. Chicago: National Educational Resources, 1978.
- Fling, S., and Manosevitz, M. Sex-typing in nursery school children's play interests. Development Psychology, 1972, 7, 146-152.
- Folb, E.A. Runnin' down some lines. Cambridge, MA: Harvard University Press, 1980.
- Foster, D. (Personal communication, 1981.)
- Fox, L.H. The effects of sex role socialization on mathematics participation and achievement. In Women and Mathematics Research: Perspectives for Change. Washington, DC: National Institute of Education, 1977.

- Fox, L.H., and Cohn, S. Sex differences in the development of precocious mathematics talent. In L.H. Fox, L. Brody, and D. Tobin (Eds.), Women and the Mathematical Mystique. Baltimore: The Johns Hopkins University Press, 1980.
- Fox, L.H., and Denham, S.A. Values and career interests of mathematically and scientifically precocious youth. In J.C. Stanley, D.P. Keating, L.H. Fox (Eds.), Mathematical Talent: Discovery, Description, and Development. Baltimore: The Johns Hopkins University Press, 1974.
- Fox, L.H., Tobin, D., and Brody, L. Sex-role socialization and achievement in mathematics. In M.A. Wittig, A.C. Petersen and M. Andrisin (Eds.), Sex-Related Differences in Cognitive Functioning. New York: Academic Press, 1979.
- Frankel, M.R., and McWilliams, H. The Profile of American Youth: Technical Sampling Report. Chicago: National Opinion Research Center, 1981.
- Frieland, S.L. Unemployment in the Urban Core. New York: Praeger, 1972.
- Garcia, E.E. Bilingualism in early childhood. Young Children, 1980, 35 (4), 52-66.
- Garcia, E.E., Maez, L., Ibanez, J., and Gonzalez, G. A national study of Spanish/English bilingualism in young Hispanic children in the United States. Paper presented at the University of Illinois-Chicago Circle Conference on Spanish Beyond the Southwest, October, 1980.
- Garfinkle, S.H. Occupations of women and Black workers, 1962-74. Monthly Labor Review, 1975, 98 (11), 25-35.
- Garland, H. and Smith, G. B. Occupational achievement motivation as function of biological sex, sex-linked personality, and occupation stereotype. Psychology of Women Quarterly, 1981, 5 (4), 568-85.
- Garron, D. Sex-linked recessive inheritance of spatial and numerical abilities and Turner's syndrome. Psychological Review, 1970, 77, 147-152.
- Gastrin, J. Det intelligenta larandets problem. Acta Academiae Aboensis, Humaniora, 1940, 13 (9).
- Gazzaniga, M.S. The Bisected Brain. New York: Appleton-Century-Crofts, 1970.
- Gemmill, L.M., Bustoz, J., and Montiel, M. Factors influencing mathematics participation of highly able Mexican-American adolescents. Final Report to the National Science Foundation on Grant No. SED 80-17768, August, 1982.
- Geshwind, N. The anatomical basis of hemispheric differentiation. In S.J. Diamond and J.G. Beaumont (Eds.), Hemispheric Function in the Human Brain. New York: Wiley, 1974.
- Ghiselli, E.E. The validity of aptitude tests in personnel selection. Personnel Psychology, 1973, 26, 461-477.
- Ghiselli, E.E. The Validity of Occupational Aptitude Tests. New York: Wiley, 1966.



- Glick, P.C. Children of divorced parents in demographic perspective. Journal of Social Issues, 1979, 35 (4), 170-82.
- Golden, M., Birns, B., Bridger, W., and Moss, A. Social class differentiation among Black children. Child Development, 1971, 42, 37-45.
- Good, T.L., Sykes, J.N., and Brody, J.E. Effects of teacher sex and student sex on classroom interaction. Journal of Educational Psychology, 1973, 65 (1), 74-87.
- Goodenough, D.R., Gandini, E., Olkin, I., Pizzamiglio, L., Thayer, D., and Witkin, H.A. A study of X chromosome linkage with field dependence and spatial visualization. Behavior Genetics, 1977, 7, 373-388.
- Gregory, M.K. Sex bias in school referrals. Journal of School Psychology, 1977, 5, 5-8.
- Guilford, J.P., and Hoepfner, R. The Analysis of Intelligence. New York: McGraw-Hill, 1971.
- Guilford, J.P., and Zimmerman, W.S. Guilford-Zimmerman Aptitude Survey. Orange, CA: Sheridan Psychological Services, 1953.
- Guilford, J.P., and Zimmerman, W.S. Some A.A.F. findings concerning aptitude factors. Occupations, 1947, 26, 154-159.
- Haberman, S. C-TAB: Analysis of Multidimensional Contingency Tables by Log-Linear Models. Chicago: National Educational Resources, 1973.
- Hale, J. The socialization of black children. Dimensions, October, 1980, 43-48.
- Hall, V.C., and Kaye, D.B. Early patterns of cognitive development. Monographs of the Society for Research in Child Development, 1980, 45, No. 2.
- Hartlage, L.C. Sex-linked inheritance of spatial ability. Perceptual and Motor Skills, 1970, 31, 610.
- Hartup, W., and Zook, E. Sex-role preferences in three- and four-year-old children. Journal of Consulting Psychology, 1960, 24, 420-426.
- Haskins, R., Walden, T., and Ramey, C.T. Teacher and student behavior in high and low ability groups. Manuscript under review.
- Hauser, R.M. Disaggregating a social-psychological model of educational attainment. In A.S. Goldberger and O.D. Duncan (Eds.), Structural Equation Models in the Social Sciences. New York: Seminar, 1973.
- Haven, E.W. Factors associated with selection of advanced academic mathematics by girls in high school. Ph.D. dissertation, University of Pennsylvania, 1971.

- Hendrickson, E.A., and White, P.O. Promax: A quick method for rotation to oblique simple structure. British Journal of Mathematical and Statistical Psychology, 1964, 17, 65-70.
- Hertzog, M.E., Birch, M.G., Thomas, A., and Mendez, O.A. Class and ethnic differences in responses of preschool children to cognitive demands. Monographs of the Society for Research in Child Development, 1968, 33, 117.
- Hess, R. Social class and ethnic influences upon socialization. In P. Mussen (Ed.), Carmichael's Manual of Child Psychology (third edition), Vol. 2. New York: Wiley, 1970.
- Hess, R., and Shipman, V. Early experience and the socialization of cognitive modes in children. Child Development, 1965, 34, 869-886.
- Hess, R., Shipman, V., Brophy, J., and Bear, R. The cognitive environments of urban preschool children: Follow-up phase. Chicago: Graduate School of Education, University of Chicago, 1969.
- Hier, D.B. and Crowley, W.F. Impaired spatial ability in men with hypogonadotropic hypogonadism. Transactions of the American Neurological Association, 1980, 105, 1-2.
- Hier, D.B., and Crowley, W.F. Spatial ability in androgen deficient men, 1982 (submitted for publication).
- Hilton, T.C. ETS study of academic prediction and growth. New Directions for Testing and Measurement, 1979, 2, 27-44.
- Hilton, T.L., and Berglund, G.W. Sex differences in mathematics achievement: A longitudinal study. Journal of Educational Research, 1974, 67, 231-237.
- Holliday, F. An investigation into selection of apprentices for the engineering industry. Occupational Psychology, 1940, 14, 69-81.
- Holliday, F. The relations between psychological test scores and subsequent proficiency of apprentices in the engineering industry. Occupational Psychology, 1943, 17, 168-185.
- Horn, J. L., and Donaldson, G. Cognitive development in adulthood. In O. Brim and J. Kagan (Eds.), Constancy and Change in Human Development. Cambridge, MA: Harvard University Press, 1980.
- Korner, M.S. Toward an understanding of achievement related conflicts in women. Journal of Social Issues, 1972, 28 (2), 157-175.
- Horowitz, F.D., and Paden, L. The effectiveness of environmental intervention programs. In B. Caldwell and H. Ricciuti (Eds.), Review of Child Development Research, Vol. 3. Chicago: University of Chicago Press, 1973.

- Hunt, E.B. Varieties of cognitive power. In L.B. Resnick (Ed.), The Nature of Intelligence. Hillsdale, NJ: Erlbaum, 1976.
- Hyde, J.S., Geiringer, E.R., and Yen, W.M. On the empirical relation between spatial ability and sex differences in other aspects of cognitive performance. Multivariate Behavioral Research, 1975, 10, 289-310.
- Hyman, H.H., Wright, C.R., Reed, J.S. The Enduring Effects of Education. Chicago: University of Chicago Press, 1975.
- Jacklin, C.N. Epilogue. In M.A. Wittig, A.C. Petersen, and M. Andrisin (Eds.), Sex-Related Differences in Cognitive Functioning. New York: Academic Press, 1979.
- Jackson, J.J. Black women in a racist society. In C.V. Willie, B. Kramer, and B. Brown (Eds.), Racism and Mental Health. Pittsburgh, PA: University of Pittsburgh Press, 1973.
- Jaeger, R.M., Linn, R.I., and Novick, M.R. A review and analysis of score calibration for the Armed Services Vocational Aptitude Battery. Appendix B in Aptitude Testing of Recruits: a Report to the House Armed Services Committee. Washington: Office of the Assistant Secretary of Defense for Manpower, Reserve Affairs and Logistics, 1980.
- Jencks, C., Bartlett, S., Corcoran, M., Crouse, J., Eaglesfield, D., Jackson, G., McClelland, K., Mueser, P., Olneck, M., Schwartz, J., Ward, S., and Williams, J. Who Gets Ahead? New York: Basic Books, 1979.
- Jencks, C., Smith, M., Acland, H., Bane, M., Cohen, D., Gintis, H., Heyns, B., and Michelson, S. Inequality. New York: Basic Books, 1972.
- Jensen, A.R. Bias in Testing. New York: Free Press, 1980.
- Jensen, A.R. Educability and Group Differences. New York: Harper and Row, 1973.
- Jensen, A.R. How much can we boost I.Q. and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.
- Jensen, A.R. The race x sex x ability interaction. In R. Canero (Ed.), Intelligence: Genetic and Environmental Influences. New York: Grune and Stratton, 1971.
- Jensen, A.R. Social class and verbal learning. In M. Deutsch, I. Katz, and A.R. Jensen (Eds.), Social Class, Race, and Psychological Development. New York: Holt, Rinehart, and Winston, Inc., 1968.
- Johnson, P.O., and Neyman, J. Tests of certain linear hypotheses and their applications to some educational problems. Statistical Research Memoirs, 1936, 1, 57-93.
- Jones, H. The environment and mental development. In L. Carmichael (Ed.), Manual of Child Psychology, Second Edition. New York: Wiley, 1954.

- Jordan, T. E. Influences on vocabulary attainment: A five-year prospective study. Child Development, 1978, 49, 1096-1106.
- Kagan, J., and Moss, A. Birth to Maturity. New York: John Wiley, 1962.
- Kagin, S., and Buriel, K. Field-dependence-independence and Mexican-American culture and education. In J. Martinez, Chicano Psychology. New York: Academic Press, 1978.
- Karnes, M.B., Wollersheim, J., Stoneburner, R., Hodgins, A., and Teska, J. An evaluation of two preschool programs for disadvantaged children: A traditional and highly structured experimental preschool. Exceptional Children, 1968, 34, 667-676.
- Kelley, T.L. Crossroads in the Mind of Man. Stanford, CA: Stanford University Press, 1928.
- Kerckhoff, A.C., and Campbell, R.T. Black-white differences in the educational attainment process. Sociology of Education, 1977, 50, 15-27.
- Kimura, D. Functional asymmetry of the brain in dichotic listening. Cortex, 1967, 3, 163-178.
- Kohn, M.L. Class and Conformity (second edition). Chicago: University of Chicago Press, 1977.
- Kolakowski, D., and Bock, R.D. A multivariate generalization of probit analysis. Biometrics, 1981, 37, 541-551.
- Kuhn, D., Nasch, S.C., and Brucken, L. Sex-role concepts of two- and three-year-olds. Child Development, 1978, 49, 445-451.
- Kulvesky, W. W. Gender differences among American youth. Paper presented at the Annual Meeting of the Southwest Sociological Association, Dallas, Texas, March 25-28, 1981.
- Labov, W. Language in the Inner-City: Studies in the Black English Vernacular. Philadelphia: University of Pennsylvania Press, 1972.
- Lesser, G., Fifer G., and Clark, D. Mental abilities of children from different social class and cultural groups. Monographs of the Society for Research in Child Development, 1965, 30, 1-115.
- Levine, M. Identification of reasons why qualified women do not pursue mathematical careers. Report to National Science Foundation, 1976.
- LeVine, R.A. Culture, personality, and socialization. In D.A. Goslin (Ed.), Handbook of Socialization Theory and Research. Chicago: Rand McNally, 1973.
- Loasa, L.M. School, occupation, culture and family: The impact of parental schooling on the parent-child relationship. Journal of Educational Psychology, 1982, 2 (6), 791-827.

- Loehlin, J.C., Sharan, S., and Jacoby, R. In pursuit of the "spatial gene": A family study. Behavior Genetics, 1978, 8, 27-41.
- Luchins, E. Women mathematicians: A contemporary reappraisal. Paper presented at the meeting of the American Association for the Advancement of Science, Boston, MA, February, 1976.
- Maccoby, E.E., and Jacklin, C.N. The Psychology of Sex Differences. Stanford, CA: Stanford University Press, 1974.
- Mackler, B. Black superstars: Getting ahead in today's America. New York: Conch Magazine Limited, 1977.
- Maier, M.H., and Fuchs, E.F., Differential Validity of the Army Aptitude Areas for Predicting Army Job Training Performance of Blacks and Whites. Technical Paper 312. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, 1978.
- Maier, M.H., and Fuchs, E.F. Effectiveness of Selection and Classification Testing. Research Report 1179. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, 1973.
- Marrett, C.B. Minority females in high school mathematics and science. Final report to the National Institute of Education on Grant No. 6-79-0110, 1981.
- Marrett, C.B. Minority females in high school mathematics and science. Madison, WI: Wisconsin Research and Development Center, University of Wisconsin, 1982.
- Martin, G.C. Test batteries for auto mechanics and apparel design. Journal of Applied Psychology, 1951, 35, 20-22.
- Matthews, W. Adding up race and sex: A study of enrollment in high school mathematics classes. Paper presented for the Program on Women, Northwestern University, Evanston, Illinois, December, 1980.
- Matthews, W. Influences on the learning and participation of minorities in mathematics. (Program Report No. 83-5.) Madison, WI: Wisconsin Center for Education Research, University of Wisconsin, 1983.
- McClendon, M.J. The occupational status attainment process of males and females. American Sociological Review, 1976, 41, 52-64.
- McCorquodale, P. Interest in science courses and careers: A comparison of Mexican-American and Anglo students. Paper presented at the Society for the Advancement of Chicanos and Native Americans in Science Conference, Albuquerque, New Mexico, November 20-22, 1980.
- McCorquodale, P. (Personal communication, 1983.)
- McFarlane, M. A study of practical ability. British Journal of Psychology, Monograph Supplement No. 8, 1925.

- McGee, M.G. Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. Psychological Bulletin, 1979, 86, 889-918.
- McGlone, J. Sex differences in human brain asymmetry: A critical survey. Behavioral and Brain Sciences, 1980, 3, 215-263.
- McMahon, I.D. Expectancy of success on sex-linked tasks. Sex Roles, 1982, 8 (9), 949-58.
- McMahon, I.D. Sex role stereotypes of cognitive task performance. Catalogue of Selected Documents in Psychology, 1976, 6, 39.
- Meece, J.L. A theoretical framework for studying students' course selection in mathematics. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, Massachusetts, April, 1980.
- Meece, J.L., Parsons, J.E., Kaczala, C.M., Goff, S.B., and Futterman, R. Sex differences in math achievement: Toward a model of academic choice. Psychological Bulletin, 1982, 91 (2), 324-48.
- Minton, C., Kagan, J., and Levine, J.A. Maternal control and obedience in two-year-old children. Child Development, 1971, 42, 1873-94.
- Montemayor, R. Children's performance in a game and their attraction to it as a function of sex-typed labels. Child Development, 1974, 45, 152-156.
- Moore, E. The effects of cultural style on black children's intelligence test achievement. Ph.D. dissertation, University of Chicago, 1980.
- Moore, E. Language behavior in the test situation and the intelligence test achievement of transracially and traditionally adopted black children. In D. Farran and L. Feagans (Eds.), Language Development of Children in Poverty. New York: Academic Press, 1981.
- Mullins, L. On women, work, and society. Freedomways, First Quarter, 1980, 15-24.
- Murillo, N. The Mexican-American family. In C. Hernandez, M. Haug, and N. Wagner (Eds.), Chicanos: Social and Psychological Perspectives. St. Louis: C. V. Mosby Company, 1976.
- Mussen, P. Early sex-role development. In D.A. Goslin (ed.), Handbook of Socialization Theory and Research. Chicago: Rand McNally, 1973.
- Nash, S.C. Sex role as a mediator of intellectual functioning. In M.A. Wittig, A.C. Petersen, and M. Andrisin, (Eds.), Sex-related Differences in Cognitive Functioning. New York: Academic Press, 1979.
- National Assessment of Educational Progress. Mathematical Applications, 1977-1978. Denver: Education Commission of the States, 1979.

- National Bureau of Standards. Bureau of Standards Tables of Bivariate Normal Distribution Function and Related Functions. Applied Mathematics Series 50. Washington: U.S. Government Printing Office, 1959.
- National Science Foundation. Women and Minorities in Science and Engineering. (NSF 82-302.) Washington, 1982.
- Nebes, R.D., and Sperry, R.W. Cerebral dominance in perception. Neuropsychologia, 1971, 9, 247.
- Nichols, P.L., and Anderson, V. Intellectual performance, race, and socioeconomic status. Social Biology, 1973, 30, 367-374.
- O'Connor, J. Structural Visualization. Boston: Human Engineering Lab, 1943.
- Ogbu, J. Minority Education and Caste. New York: Academic Press, 1978.
- Ornstein, R.E. The Psychology of Consciousness. San Francisco: Freeman, 1973.
- Pallas, A.M., and Alexander, K.L. Sex differences in quantitative SAT performance: New evidence on the differential coursework hypothesis. American Educational Research Journal, 1983, 20, 165-182.
- Park, J., Johnson, R.C., et al. Parent-offspring resemblances for specific cognitive abilities in Korea. Behavior Genetics, 1978, 8, 43-52.
- Parsons, J.E., Adler, T.F. and Kaczala, C.M. Socialization of achievement attitudes and beliefs: Parental influences. Child Development, 1982, 53, 310-321.
- Parsons, J.E., Kaczala, C.M., and Meece, J.L. Socialization of achievement attitudes and beliefs: Classroom experiences. Child Development, 1982, 53, 322-339.
- Pedersen, E., Faucher, T.A., and Eaton, W.W. A new perspective on the effects of first-grade teachers on children's subsequent adult status. Harvard Educational Review, 1978, 43 (1), 1-31.
- Pedraza-Bailey, S. Cubans and Mexicans in the United States: The functions of political and economic migration. Cuban Studies/Estudios Cubanos, 1982, 11:2/12:1, 79-97.
- Petersen, A.C. Hormones and cognitive functioning in normal development. In M.A. Wittig, A.C. Petersen, and M. Andrisin, (Eds.), Sex-Related Differences in Cognitive Functioning. New York: Academic Press, 1979.
- Petersen, A.C. (Personal communication, 1981).
- Petersen, A.C. Physical androgyny and cognitive functioning in adolescence. Developmental Psychology, 1976, 12, 524-533.

- Porter, J.N. Race, socialization, and mobility in educational and early occupational attainment. American Sociological Review, 1974, 39, 303-316.
- Portes, A., McLeod, S.A., and Parker, R.N. Immigrant aspirations. Sociology of Education, 1978, 51, 241-260.
- Portes, A., and Wilson, K.L. Black-White differences in educational attainment. American Sociological Review, 1976, 41, 414-431.
- Prediger, D.J., McClure, G.I., and Noeth, R.J. Promoting the Exploration of Personality Relevant Career Options in Science and Technology. Report to the National Science Foundation, 1976.
- Rabban, M. Sex role identification of young children in two diverse social groups. Genetic Psychology Monographs, 1950, 42, 51-158.
- Ramey, C.T., and Haskins, R. Early education, intellectual development, and school performance: A reply to Arthur Jensen and J. McVicker Hunt. Intelligence, 1981, 5, 41-48.
- Ramirez, M. Cognitive Styles and Cultural Democracy in Education. Social Science Quarterly, 1973, 53, 895-904.
- Rasch, G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press, 1980. (Expanded edition.)
- Rist, R.C. The Urban School as a Factory for Failure. Cambridge, Massachusetts: M.I.T. Press, 1973.
- Robovits, P. Early experience and the achieving orientations of American middle-class girls. In M. Maehr and W. Stallings (Eds.), Culture, Child, and School. Monterey, Calif.: Brooks/Cole, 1975.
- Rosenberg, M., and Rosenberg, F. The occupational self: A developmental study. Paper presented at the Self-Concept Symposium, Boston, September 28 - October 1, 1978.
- Roy, S.N. Some Aspects of Multivariate Analysis. New York: Wiley, 1957.
- Scannell, D.P. Test of Academic Progress, Manual, Form S. Boston: Houghton Mifflin, 1972.
- Scarr, S. Commentary on V.C. Hall and D.B. Kaye, Early patterns of cognitive development. Monographs of the Society for Research in Child Development, 1980, 45, No. 2.
- Scarr, S. Genetics and the development of intelligence. In F.D. Horowitz (Ed.), Review of Child Development Research, Vol. 4. Chicago: University of Chicago Press, 1975.
- Scarr, S. Race, social class, and I.Q. Science, 1971, 174, 1285-1292.



- Scarr, S., Pakstis, A., Katz, S., and Barker, W. The absence of a relationship between degree of White ancestry and intellectual skills within the Black population. Human Genetics, 1977, 39, 69-86.
- Scarr, S. and Weinberg, R. I.Q. test performance of black children adopted by white families. American Psychologist, 1976, 31, 726-739.
- Schiff, M., Duyme, M., Dumaret, A. and Tomkiewicz, S. How much could we boost scholastic achievement and I.Q. scores?: A direct answer from a French adoption study. Cognition, 1982, 12, 165-96.
- Seeley, L.C., Fischel, M.A., and Hicks, J.M. Development of the ASVAB, Forms 2 and 3. Technical Paper 289. Alexandria, VA: Army Research Institute.
- Segal, N.L. Cooperation, competition, and altruism within twin-sets: A reappraisal. Ph.D. dissertation, University of Chicago, 1982.
- Sells, L.W. High school mathematics as a critical filter in the jobs market. In R.T. Thomas (Ed.), Developing Opportunities for Minorities in Graduate Education. Proceedings of the Conference on Minority Graduate Education, University of California, Berkeley, May 11-12, 1973.
- Sells, L.W. The mathematics filter and the education of women and minorities. In L.H. Fox, L. Brody, and D. Tobin (Eds.), Women and the Mathematical Mystique. Baltimore: The Johns Hopkins University Press, 1980.
- Serbin, L., O'Leary, K., Kent, R., and Tonick, I. A comparison of teacher response to the pre-academic and problem behavior of boys and girls. Child Development, 1973, 44, 796-804.
- Sewell, T., Farley, F.H., Manni, J., and Hunt, P. Motivation, social reinforcement, and intelligence as predictors of academic achievement in Black adolescents. Adolescence, 1982, 18 (67), 647-56.
- Sewell, T.E. and Martin, R.P. Racial patterns of occupational choice in adolescents. Psychology in the School, 1976, 13, (3), 326-333.
- Sewell, W.H., and Hauser, R.M. Education, Occupation, and Earnings. New York: Academic Press, 1975.
- Sheatsley, P.B. The Profile of American Youth: Pretest Report. Chicago: National Opinion Research Center, 1980.
- Shenkel, K.F., Leedy, H.B., Rosenberg, N., and Mundy, J.P. Evaluation of the Puerto Rican screening test (ECFA) against success in training. Technical Research Report PRB1097. Washington, D.C.: Personnel Research Branch, The Adjutant General's Office, 1957.
- Shepard, R.N., and Metzler, J. Mental rotation of three-dimensional objects. Science, 1971, 171, 701-703.

- Sherman, J. Girls' and boys' enrollments in theoretical math courses: A longitudinal study. Psychology of Women Quarterly, 1981, 5 (Supplement), 681-89.
- Sherman, J. Mathematics, the critical filter: A look at some residues. Psychology of Women Quarterly, 1982, 6 (4), 428-45.
- Sherman, J.A. Problem of sex differences in space perception and aspects of intellectual functioning. Psychological Review, 1967, 74, 290-299.
- Sherman, J.M., and Fennema, E. The study of mathematics among high school girls and boys: Related factors. American Educational Research Journal, 1977, 14 (2), 159-168.
- Sims, W.H. Interim results of an examination of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 6 and 7 for the classification of Marine Corps recruits. CNA Report 78-8031. Alexandria, VA: Center for Naval Analysis, 1978.
- Sitkei, E., and Meyers, C. Comparative structure of intellect in middle and lower class four-year-olds of two ethnic groups. Developmental Psychology, 1969, 1, 592-604.
- Slater, P. Tests for selecting secondary and technical school children. Occupational Psychology, 1941, 15, 10.
- Slaughter, D.T. Education and the family. In B. Bloom and C. Mathias (Eds.), The Schools, Schools of Education, and Major Changes Affecting Education: The Crises of Our Time. (In preparation)
- Slaughter, D.T. On becoming an Afro-American woman. School Review, 1972, 80, 299-318.
- Smith, I.M. Spatial Ability. San Diego: Robert R. Knapp, 1964.
- Solano, C.H. Teacher and people stereotypes of gifted boys and girls. Paper presented at the 1976 meeting of the American Psychological Association, Washington, D.C., September, 1976.
- Solomon, L.C., and Taubman, P. (Eds.) Does College Matter? New York: Academic Press, 1973.
- Sperry, R.W. Perception in the absence of neocortical commissures. In Perception and Its Disorders, Res. Publ. A.R.N.M.D., vol. 48, The Association for Research in Nervous and Mental Disease.
- Stafford, R.E. Hereditary and environmental components of quantitative reasoning. Review of Educational Research, 1972, 42, 183-201.
- Stafford, R.E. Sex differences in spatial visualization as evidence of sex-linked inheritance. Perceptual and Motor Skills, 1961, 13, 428.
- Stein, A.H. The effects of sex-role standards for achievement and sex-role preference on three determinants of achievement motivation. Developmental Psychology, 1971, 4, 219-231.

- Stein, A.H., Pohly, S.R., and Mueller, E. The influence of masculine, feminine, and neutral tasks on children's achievement behavior, expectancies of success, and attainment values. Child Development, 1971, 42, 195-207.
- Stein, A.H., and Smithell, J. Age and sex differences in children's sex role standards about achievement. Developmental Psychology, 1969, 1, 252-259.
- Streissguth, A.P., and Bee, H.L. Mother-child interactions and cognitive development in children. In W.W. Hartup (Ed.), The Young Child: Reviews of Research, Vol. 2. Washington, D.C.: Association for the Education of Young Children, 1972.
- Super, D.E., and Crites, J.O. Appraising Vocational Fitness. New York: Harper and Row, 1962.
- Swanson, L. Armed Services Vocational Aptitude Battery, Forms 6 and 7: Validation Against School Performance in Navy Enlisted Schools (July 1976-February 1978). NPRDC Technical Report 80-1. San Diego, Calif.: Navy Personnel and Training Research Laboratory, 1979.
- Sympson, J.B., and Weiss, D.J. Predictive validity of conventional and adaptive tests in an Air Force training environment. Final report on Contract No. F33615-77-C-0061. Brooks Air Force Base, TX: Air Force Human Resources Laboratory, 1981.
- Taylor, H.C. and Russell, J.T. The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. Journal of Applied Psychology, 1939, 28, 565-578.
- Thoday, J.M. Limitations of genetic comparisons of populations. Journal of Biosocial Science, 1969, Supplement, 3-14.
- Thomas, G.E. Race and sex differences and similarities in the process of college entry. Higher Education, 1980, 9, 179-202.
- Thomas, G.E., Alexander, K.L., and Eckland, B.K. Access to higher education: The importance of race, sex, social class, and academic credentials. School Review, 1979, 87, 133-56.
- Thurstone, L.L., and Thurstone, T.G. Factorial studies of intelligence. Psychometric Monographs, no. 2. Chicago, University of Chicago Press, 1941.
- Thurstone, L.L., and Thurstone, T.G. The Primary Mental Abilities. Chicago: Science Research Associates, 1947.
- Tipton, R. Attitudes towards women's roles in society and vocational interests. Journal of Vocational Behavior, 1976, 8, 155-165.
- Tobias, S. Math anxiety: Why is a smart girl like you counting on your fingers. Ms. Magazine, 1976, 92, 56-59.

- Tobin, D., and Fox, L.H. Career interests and career education: A key to change.  
In L.H. Fox, L. Brody, and D. Tobin (Eds.), Women and the Mathematical Mystique. Baltimore: The Johns Hopkins University Press, 1980.
- Treiman, D.J., and Terrell, K. Sex and the process of status attainment: A comparison of working women and men. American Sociological Review, 1975, 40, 174-200.
- Trotman, F. Race, I.Q., and the middle class. Journal of Educational Psychology, 1977, 69 (3), 266-273.
- Tulkin, S., and Covitz, F. Mother-infant interaction and intellectual functioning at age 6. Paper presented at the Biannual Meetings of the Society for Research in Child Development, Denver, Colorado, 1975.
- Tulkin, S., and Kagan, J. Mother-infant interaction in the first year of life. Child Development, 1972, 43, 31-41.
- Turner, H. (Personal communication, 1983.)
- United States Bureau of the Census. Educational Attainment in the United States: March 1979 and 1978. Current Population Reports, Series P-20, no. 356, August. 1980.
- United States Bureau of the Census. Statistical Abstract of the U.S. Washington, D.C.: U.S. Government Printing Office, 1980.
- United States Commission on Civil Rights. Ethnic isolation of Mexican-Americans in public schools of the Southwest. Mexican-American Study Project. Washington, D.C.: U.S. Government Printing Office, 1971.
- Vaidya, S., and Chansky, N. Cognitive development and cognitive style in mathematics achievement. Journal of Educational Psychology, 1980, 72, 326-330.
- Valentine, L.D. Prediction of Air Force technical training success from ASVAB and educational background. AFHRL-TR-77-18. Brooks Air Force Base, Texas, Air Force Human Resources Laboratory, 1977.
- Vandenberg, S.G. Assortive mating, or who marries whom? Behavior Genetics, 1972, 2, 127-157.
- Vandenberg, S.G. Hereditary factors in psychological variables in man, with special emphasis on cognition. In J.N. Spuhler (Ed.), Genetic diversity and behavior. Chicago: Aldine, 1967.
- Vandenberg, S.G. Sources of variance in performance on spatial tests. In J. Eliot and N.J. Salkind (Eds.), Children's Spatial Development. Springfield, IL: Charles C. Thomas, 1975.
- Vitola, B.M., Mullins, C.J., and Croll, P.R. Validity of Armed Services Vocational Aptitude Battery Form 1, to predict technical school success. (AFHRL-TR Report 73-7.) Brooks Air Force Base, TX: Air Force

- Human Resources Laboratory, 1973.
- Waber, D.F. Sex differences in mental abilities, hemisphere lateralization, and rate of physical growth at adolescence. Developmental Psychology, 1977, 13, 28-38.
- Walker, J.T., Krasnoff, A.G., and Peoco, D. Visual spatial perception in adolescents and their parents: The x-linked recessive hypothesis. Behavior Genetics, 1981, 11, 403-413.
- Weiner, B. Achievement Motivation and Attribution Theory. Morristown, New Jersey: General Learning Press, 1974.
- Werdelin, I. Geometrical Ability and Space Factors in Boys and Girls. Lund, Sweden: University of Lund, 1961.
- Wilfong, H.D. ASVAB: Technical supplement to the high school counselor's guide. Fort Sheridan, IL: Directorate of Testing, United States Military Enlistment Processing Command, 1980.
- Williams, J.H., and Whitney, D. Vocational interests of minority disadvantaged students: Are they different? The Negro Educational Review, 1978, 29 (2), 97-103.
- Williams, T. Abilities and environments. In W.H. Sewell, R.M. Hauser, and D.L. Featherman (Eds.), Schooling and Achievement in American Society. New York: Academic Press, 1976.
- Williams, T. Competence dimensions of family environments. Paper presented at the Annual Meetings of the American Educational Research Association, Chicago, Illinois, 1974.
- Willerman, L. Social aspects of minimal brain dysfunction. In F.F. de la Cruz, B.H. Fox, R.H. Roberts, and G. Tarjan (Eds.), Minimal brain dysfunction. Annals of the New York Academy of Sciences, 1973, 205, 164-172.
- Willerman, L., and Stafford, R.E. Maternal effects on intellectual functioning. Behavior Genetics, 1972, 2 (4), 321-325.
- Wilson, K.L., and Portes, A. Immigrant enclaves: An analysis of labor market experiences of Cubans in Miami. American Journal of Sociology, 1980, 86 (2), 295-319.
- Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1971.
- Witkin, H., Dyk, R., Fatterson, H., Goodenough, D., and Karp, S. Psychological Differentiation. New York: Wiley, 1962.
- Witty, P., Garfield, S., Brink, W. A comparison of vocational interests of Negro and White high school students. Journal of Educational Psychology, 1941, 32, 124-132.

- Wolf, R. The measurement of environments. In A. Anastasi (Ed.), Testing Problems in Perspective. New York: American Council on Education, 1964.
- Wolfeat, P.L., Pedro, J.D., Becker, A., and Fennema, F. Sex differences in high school students' attribution of performance in mathematics. Journal of Research in Mathematics Education, 1980, 357-366.
- Yen, W.M. Sex-linked major-gene influences on selected types of spatial performance. Behavior Genetics, 1975, 5 (3), 281-298.
- Zuckerman, D. M. and Sayre, D. H. Cultural sex-role expectations and children's sex-role concepts. Sex Roles, 1982, 8 (8), 853-864.